

Office of Research and Development

**Engaging Undergraduate Psychology Students with Research
Methods and with the Process of Conducting Research**

Peter J Allen

**This thesis is presented for the Degree of
Doctor of Philosophy
of
Curtin University**

October 2016

Declaration

To the best of my knowledge and belief this thesis contains no material previously published by any other person except where due acknowledgment has been made.

This thesis contains no material which has been accepted for the award of any other degree or diploma in any university.

Human Ethics: The research presented and reported in this thesis was conducted in accordance with the National Health and Medical Research Council National Statement on Ethical Conduct in Human Research (2007) – updated March 2014. The proposed research studies received human research ethics approval from the Curtin University Human Research Ethics Committee (EC00262), Approval Numbers: Psych 2009 02, Psych and SP 2010 18, Psych and SP 2010 33, Psych and SP 2011 35, Psych and SP 2012 01, Psych and SP 2012 02, Psych and SP 2012 45, Psych and SP 2013 86.

Signature:

A handwritten signature in black ink, appearing to be 'E. Au', written over a horizontal line.

Date: 7 October 2016

Acknowledgements

The enormous contribution that Lynne Roberts has made to this thesis should be immediately evident. There are nine papers contained herein; Lynne co-authored eight of them.

I'd also like to acknowledge my other co-authors (Frank Baughman, Kate Dorozenko, James Finlay, Natalie Loxton, Amanda Lourenco, Dirk Van Rooy, and Adam Rock), 1000+ participants, and the many peer-reviewers and journal editors who made each of these papers a reality.

Thank you to my colleagues, past and present - including particularly Jan Piek, Leigh Smith, Lyndall Steed, Clare Pollock, Ros Morrow, Brian Bishop, Mara Blosfelds, Natalie Gasson, and Janet Beilby - for your support, mentorship and advice.

And finally, to anyone who's ever asked, "is your PhD finished yet?" You can block up now. It's done.

Abstract

A strong understanding of quantitative research methods is a pre-requisite to psychological literacy and evidence-based practice in psychology. Quantitative research methods are also an area of weakness for many psychology students. Furthermore, many students have relatively little interest in reading and conducting research, hold negative attitudes toward research methods, struggle to see the relevance or utility of methods and statistics courses, and experience high levels of statistics anxiety. Consequently, efforts have been undertaken to reform traditional research methods and statistics pedagogy, with the objective of making these subjects more applied, relevant and engaging for students. Many of these reforms are based on active learning principles, and the idea that, as much as is practicable, students should be ‘doing’ research, rather than merely reading about it, or listening to instructors talking about it. In an undergraduate psychology degree, ‘doing research’ can manifest in multiple activities, of which the current thesis focuses on three: (1) participating in authentic research; (2) working with authentic data; and (3) conducting an original research project.

The first two papers herein focus on understanding and quantifying undergraduate psychology students’ perspectives on the educational value of participating in authentic research, which is a ‘rite of passage’ in most research active schools of psychology. The third describes the development and evaluation of an active learning exercise in which students participated in a class experiment, then analysed the data it generated. Papers 4-7 address issues arising from the supervision of final year dissertations projects, including the quality of student collected data, and the ethics of surveying online. Finally, paper 8 explores the difficulties faced by students (but not ‘experts’) when required to identify statistical tests and procedures appropriate to their research questions and hypotheses, while paper 9 describes the development of a mobile application specifically developed to support this process.

Combined with the exegesis that precedes them, the nine papers in this thesis offer a range of insights into, and strategies that promote the engagement of undergraduate psychology students with research methods, and with the process of conducting research.

List of Publications Included in this Thesis

- Roberts, L. D., & **Allen, P. J.** (2012). Student perspectives on the value of research participation. In S. McCarthy, K. L. Dickson, J. Cranney, A. Trapp, & V. Karandashev (Eds.), *Teaching psychology around the world* (vol 3, pp. 198-211). Newcastle, England: Cambridge Scholars.
- Roberts, L. D., & **Allen, P. J.** (2013). A brief measure of student perceptions of the educational value of research participation. *Australian Journal of Psychology*, *65*, 22-29. doi:10.1111/ajpy.12007
- Allen, P. J.**, & Baughman, F. D. (2016). Active learning in research methods classes is associated with higher knowledge and confidence, though not evaluations or satisfaction. *Frontiers in Psychology*, *7*, Article 279. doi:10.3389/fpsyg.2016.00279
- Allen, P. J.**, Lourenco, A., & Roberts, L. D. (2016). Detecting duplication in students' research data: A method and illustration. *Ethics & Behavior*, *26*, 300-311. doi:10.1080/10508422.2015.1019070
- Allen, P. J.**, & Roberts, L.D. (2016). The impact of academic sponsorship on Web survey dropout and item non-response. *First Monday*, *21*(2). doi:10.5210/fm.v21i2.6144
- Allen, P. J.**, & Roberts, L. D. (2010). The ethics of outsourcing online survey research. *International Journal of Technoethics*, *1*, 35-48. doi:10.4018/jte.2010070104
- Roberts, L. D., & **Allen, P. J.** (2015). Exploring ethical issues associated with using online surveys in educational research. *Educational Research and Evaluation*, *21*, 95-108. doi:10.1080/13803611.2015.1024421
- Allen, P. J.**, Dorozenko, K. P., & Roberts, L. D. (2016) Difficult decisions: A qualitative exploration of the statistical decision making process from the perspectives of psychology students and academics. *Frontiers in Psychology*, *7*, Article 188. doi:10.3389/fpsyg.2016.00188
- Allen, P. J.**, Roberts, L. D., Baughman, F. D., Loxton, N. J., Van Rooy, D., Rock, A. J., & Finlay, J. (2016). Introducing StatHand: A cross-platform mobile application to support students' statistical decision making. *Frontiers in Psychology*, *7*, Article 288. doi:10.3389/fpsyg.2016.00288

I warrant that I have obtained, where necessary, permission from the copyright owners to use any third-party copyright material reproduced in the thesis (e.g. questionnaires, artwork, unpublished letters), or to use any of my own published work (e.g. journal articles) in which the copyright is held by another party (e.g. publisher, co-author).

Statement of Author Contributions

Co-author statements declaring and endorsing the candidate's contributions to each paper included in this thesis can be found in Appendix A.

Table of Contents

Declaration.....	i
Acknowledgements.....	ii
Abstract.....	iii
List of Publications Included in this Thesis.....	iv
Statement of Author Contributions.....	vi
Table of Contents.....	vii
Exegesis	1
Introduction.....	2
Participation in Authentic Research.....	4
Working with Authentic Data.....	7
Conducting Original Research.....	9
Challenges in Conducting Research	15
Summary of Thesis Objective and Aims	18
Conclusion	22
References.....	26
Paper 1: Student Perspectives on the Value of Research Participation	51
Paper 2: A Brief Measure of Student Perceptions of the Educational Value of Research Participation	74
Paper 3: Active Learning in Research Methods Classes is Associated with Higher Knowledge and Confidence, though not Evaluations or Satisfaction.....	99
Paper 4: Detecting Duplication in Students' Research Data: A Method and Illustration.....	107
Paper 5: The Impact of Academic Sponsorship on Web Survey Dropout and Item Non-Response.....	135
Paper 6: The Ethics of Outsourcing Online Survey Research.....	144

Paper 7: Exploring Ethical Issues Associated with Using Online Surveys in Educational Research.....	159
Paper 8: Difficult Decisions: A Qualitative Exploration of the Statistical Decision Making Process from the Perspectives of Psychology Students and Academics....	189
Paper 9: Introducing StatHand: A Cross-Platform Mobile Application to Support Students' Statistical Decision Making.....	204
Appendix A: Author Contribution Statements	215
Appendix B: Copyright Permissions	225
Appendix C: List of Additional Peer-Reviewed Publications Since 2008	234

Exegesis

Introduction

This history and progress of modern psychology rests primarily on quantitative foundations (Benjamin, 2014; Brysbaert & Rastie, 2013). The vast majority of contemporary published psychological research is quantitative (Kidd, 2002; Marchel & Owens, 2007; Munley, Anderson, Briggs, Devries, Forshee, & Whisner, 2002; Rennie, Watson, & Monteiro, 2002), and the ability to read, comprehend, critically evaluate, and apply the findings of this body of work is a prerequisite to psychological literacy (McGovern et al., 2010; Cranney & Dunn, 2011; Cranney, Morris, & Bolwood, 2015; Roberts, Heritage, & Gasson, 2015), and evidence based practice in psychology (American Psychological Association Presidential Task Force on Evidence Based Practice, 2006). It is for these reasons that quantitative research methods hold a prominent position in the undergraduate psychology curriculum in Australia (Lipp et al., 2007; P. Wilson & Provost, 2006), the United States (D. Dunn et al., 2010; W. Messer, Griggs, & Jackson, 1999; Perlman & McCann, 1999a, 1999b; Norcross et al., 2016; Stoloff et al., 2010; Stoloff, Curtis, Rodgers, Brewster, & McCarthy, 2012), the United Kingdom (Field, 2010), and elsewhere (e.g., Gines, 2006; Sumer, 2016).

The centrality of research methods to the discipline is also clearly reflected in the learning goals, standards, and graduate attributes or competencies specified by psychology course accreditation agencies worldwide. For example, of the six graduate attributes for an Australian undergraduate psychology course developed by Cranney and colleagues (2009) and later adopted by the Australian Psychology Accreditation Council (APAC, 2010) for inclusion in the *Accreditation Standards for Psychology Courses*, at least two require a solid and flexible understanding of research methods and statistics. The first of these, “research methods in psychology”, specifies the ability to “describe, apply and evaluate the different research methods used by psychologists”, and “design and conduct basic studies to address psychological questions” (APAC, 2010, p. 40). The second, “critical thinking skills”, states that graduates of undergraduate psychology courses must be able to “apply knowledge of the scientific method in thinking about problems related to behaviour and mental processes” (APAC, 2010, p. 41). Although the APAC standards are undergoing review at the time of writing, the current consultation draft (dated June 2016) indicates that research methods and statistics will continue to play a significant role in undergraduate

psychology training in Australia. For example, the draft standards state that, amongst other competencies, graduates of four-year undergraduate psychology degrees 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 American Psychological Association (APA) Board of Educational Affairs Task Force on Psychology Major Competencies (2013; see also APA, 2016) have specified five learning goals for an undergraduate psychology degree. The second of these 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” (APA Board of Educational Affairs Task Force on Psychology Major Competencies, 2013, p. 15). Similarly, in the United Kingdom, the Quality Assurance Agency for Higher Education (2010) and the British Psychological Society (2015) both state that an undergraduate psychology degree should equip students with the ability to “generate and explore research questions”, “carry out empirical studies involving a variety of methods of data collection”, “analyse data”, “present and evaluate research findings”, and “employ evidence based reasoning”.

Despite their role in the progress of psychological science, their place in the undergraduate curriculum, and the regularity with which they appear in accreditation standards worldwide, research methods and (particularly) statistics are areas of weakness for many students (Garfield & Ahlgren, 1988; Garfield & Ben-Zvi, 2007; Murtonen, 2015; Murtonen & Lehtinen, 2003; Murtonen, Olkinuora, Tynjala, & Lehtinen, 2008). Furthermore, many students have relatively little interest in reading and conducting research (Rottinghaus, Gaffey, Borgen, & Ralston, 2006; Vittengl et al., 2004), hold negative attitudes towards research methods (Addison, Stowell, & Reab, 2015; Murtonen, 2005; Sizemore & Lewandowski, 2009), fail to see the future relevance or utility of methods and statistics courses (Ciarocco, Lewandowski, & Van Volkom, 2013; Murtonen et al., 2008), and experience high levels of statistics anxiety (Hanna, Shevlin, & Dempster, 2008; Macher, Papousek, Ruggeri, & Paechter, 2015; Onwuegbuzie & Wilson, 2003). For these reasons and more, research methods and

statistics are generally regarded as challenging subjects to teach (Connors, McCown, & Roskos-Ewoldsen, 1998; P. Dunn, Carey, Richardson, & McDonald, 2016; Saville, 2015; S. Wilson, 2013). Consequently, there have been many calls for reforms to traditional research methods and statistics pedagogy, with the objective of making these topics more applied, relevant, and engaging for students (e.g., Garfield, Hogg, Schau, & Whittinghill, 2002; Hogg, 1991; Lovett & Greenhouse, 2000). One set of reforms that have met with success are those focused on active learning, and the idea that, as much as is practicable, students should be ‘doing’ research, rather than merely reading about it, or listening to instructors talking about it (D. Dunn, 2010, 2015; Earley, 2014; Harlow, 2013; Gurung et al., 2016).

In the context of an undergraduate psychology degree, ‘doing research’ can manifest in multiple activities. These activities are most commonly deployed in dedicated methods and statistics classes, though they are also incorporated with some degree of regularity in ‘topic’ classes as well (Perlman & McCann, 2005). They include, but are not limited to, (a) engaging in class activities designed to illustrate methodological and statistical concepts; (b) participating in authentic research; (c) working with authentic data, collected by students, either by testing themselves, each other, or participants external to the unit in which they are enrolled; and (d) sharing or taking responsibility for conducting an original research project. This thesis is focused on research and initiatives related to the latter three types of active learning experience. Its overarching objective is:

To probe issues and strategies linked to the engagement of undergraduate psychology students with research methods, and the process of conducting research.

Participation in Authentic Research

The recruitment of undergraduate students as participants in honours, postgraduate, and faculty research has long been a part of the culture of psychological science (e.g., McNemar, 1946). The majority of psychological research is based on data collected from undergraduate students (Arnett, 2008; Higbee, Millard, & Folkman, 1982; Korn, 1999; Wintre, North, & Sugar, 2001), who are typically members of subject or participant pools, and required to participate in research (or complete other ‘equivalent’ activities) as a course requirement (Coulson, 1999; Diamond & Reidpath, 1992; Sieber & Saks, 1989). Whilst subject to a number of

(primarily ethical) critiques (e.g., Bartholomay & Sifers, 2016; Dalziel, 1996; Sieber, 1999; Tabachnick, Keith-Spiegel, & Pope, 1991), participant pools are typically justified on the grounds that they are mutually beneficial. They provide faculty with ready access to research participants, whilst providing students with insights into the research process that would be difficult or impossible to acquire in a typical research methods classroom (Sieber & Saks, 1989). A limited body of existing research suggests that some, though not all students see research participation as having some educational value (Cromer, Reynolds, & Johnson, 2013; Davis & Fernald, 1975; Landrum & Chastain, 1995; Leak, 1981; Moreland, 1999; Trafimow, Madson, & Gwizdowski, 2006; VanWormer, Jordan, & Dlalock, 2014), with the nature and extent of self-reported benefits influenced by a variety of factors, including the quality of debriefing provided (King, 1970), amount of previous participation experience (Nimmer & Handelsman, 1992), and total hours of participation required to meet specified course targets (Miles, Cromer, & Narayan, 2015). Most of this research has been quantitative (though see Brody, Gluck, & Aragon, 2000; Moyer & Franklin, 2011), with data typically limited to responses to a handful of survey items. To redress this imbalance, the first aim of this thesis was:

To develop a rich, qualitative account of undergraduate psychology students' perspectives on the educational value of research participation.

This aim was addressed with Roberts and Allen (2012; paper 1), in which we thematically analysed students' answers to the question, "you've been invited to participate in a number of research projects this semester; what have you learned from this experience?" At the time the research was conducted (circa 2009/10), the School of Psychology and Speech Pathology, Curtin University, did not have a participant pool (though there was active faculty discussion around establishing one). Lynne Roberts and I both coordinated research methods units, and routinely encouraged our students to participate in the research of their more senior colleagues (e.g., Honours and PhD students). The question above, which students were informed would be included in their final exams and, importantly, could be answered regardless of the extent and nature of a student's research participation, reflected one method that we used to encourage participation. The results reported in Roberts and Allen (2012) offer a rich, nuanced and contextually sensitive account of the perceived value of the research participation experience from the perspective of a near complete 'population'

of students. The overarching theme to emerge from the data was that research participation provides increased insight into the research process. This theme was expressed in several sub-themes, which captured students' developing awareness of the diversity of psychological research; the nature and complexity of 'real' research; the roles of the researcher and participant in the research dynamic; multiple design issues; and their own possible future selves as researchers. From these findings, we concluded that most students in our sample appeared to have valued the opportunity to participate in research, and that doing so afforded them a variety of educational benefits. However, we argued that greater gains could be realised through tighter integration of research participation and the teaching of research methods, and proposed several strategies toward achieving this. These strategies included using research projects in which students have recently participated as the context for exploring methodological and statistical concepts in class, and inviting the researchers themselves to brief students on their methods and findings at appropriate points throughout the syllabus.

To evaluate the effectiveness of the types of strategies we proposed in Roberts and Allen (2012), as well as track students' perceptions of the educational value of research participation over time and across circumstances, a brief, reliable, and content valid measure of these perceptions is required. Historically, when researchers have measured students' perceptions of the educational value of research participation, they have tended to do so using either single item measures, or scales with unknown or un/under-reported psychometric properties that are not clearly grounded in the substantive domain they seek to assess (e.g., King, 1970; Landrum & Chastain, 1995; Leak, 1981; Trafimow et al., 2006; VanWormer et al., 2014). Redressing these deficits was the second aim of the current thesis:

To develop a brief, reliable and valid measure of the perceived educational value of research participation, grounded in qualitative perspectives of student research participants, as captured by Roberts and Allen (2012).

This aim was met with Roberts and Allen (2013; paper 2), in which we described the development and validation of the seven-item *Student Perceptions of the Educational Value of Research Participation Scale* (SPEVRPS). Following the publication of the SPEVRPS in early 2013, Miles and colleagues (2015) developed and published the *Human Subject Pool Attitude Scale* (HSP-AS). Like items on the

SPEVRPS, items on the four sub-scales of the HSP-AS were developed from thematic analysis of students' qualitative reflections on the research experience, a strategy recommended (Gehlbach & Brinkworth, 2011) and frequently used (e.g., Fredricks et al., 2016; Nichter, Nichter, Thompson, Shiffman, & Moscicki, 2002; Rowan & Wulff, 2007) in the earlier stages of scale development as a means of ensuring content validity. Furthermore, the final 12-item HSP-AS subscale, *Educational*, shares substantial content overlap with the SPEVRPS. For example, HSP-AS item 32, "participation gives students an idea of how to conduct their own research", is conceptually similar to SPEVRPS item 1, "increased my knowledge of how research is conducted". Similarly, "I learned about the different kinds of psychological research" (HSP-AS Item 33) and "increased my knowledge about the range of research conducted in my university" (SPEVRPS item 2) appear to overlap, along with "research participation adds to what we learn in class" (HSP-AS item 35) and "I have been able to put what we have learned in class into context" (SPEVRPS item 6). Despite these similarities, the HSP-AS appears to have been developed and published without any awareness of Roberts and Allen (2013), and thus opportunities for further construct validation of the SPEVRPS, and initial construct validation of the HSP-AS were missed. At the time of writing, there have been no psychometric properties published for the HSP-AS beyond internal consistency coefficients (Miles et al., 2015).

Working with Authentic Data

Beyond learning about the conduct of research by participating in research, research suggests that research methods and statistics can be effectively taught via class exercises that engage students as researchers, who are charged with systematically measuring then analysing their own behaviour, or the behaviour of each other (e.g., Hamilton & Geraci, 2004; Morgan, 2009; Neumann, Hood, & Neumann, 2013; Neumann, Neumann, & Hood, 2010; Stedman, 1993; Thompson, 1994). Assuming responsibility for collecting data provides students with opportunities to organically engage with a range of design related issues (e.g., sampling, ethics, experimenter bias, measurement fidelity etc.). Subsequently analysing this personally meaningful data provides opportunities to reflect on the relationships between design and analysis; use design features (which are relatively easy to remember) as mnemonics to aid the recall of statistical techniques (which are comparatively harder

for most students to remember); and also, depending on the nature of the research questions posed, learn something interesting about psychology to boot (Singer & Willett, 1990). Extant research indicates that students enjoy using real, class-generated data. Furthermore, they tend to self-report that the use of such data helps them understand key methodological and statistical concepts, and endorse their use in future classes (Hamilton & Geraci, 2004; Lipsitz, 2000; Neumann et al., 2010; Ragozzine, 2002; Stedman, 1993; Thompson, 1994). Very few studies have attempted to objectively assess the learning resulting from the use of student generated data in class activities, and those that have lack internal validity. For example, when Morgan (2009) sought to assess the impact of a class project on knowledge of single-case research design and statistical process control, he did so with a pre-experimental one-group pretest-posttest design (Shadish, Cook, & Campbell, 2002), in which the same outcome measure (a 10-item multiple choice quiz) was used for both the pre- and post-tests.

The third aim of the current thesis was motivated by the gains students self-report after working with class-generated data, concerns over the paucity of objective data supporting their use, and a desire to increase student engagement in one of my undergraduate research methods and statistics units:

To (a) develop a class exercise in which students participate in an experiment, engage in class discussion around its methods, then use class-generated data to practice various data handling and statistical procedures; and (b) evaluate the aforementioned exercise in terms of (i) its subjective appeal to students; and (ii) its pedagogic effectiveness.

This aim was achieved with Allen and Baughman (2016; paper 3). The experiment was computer based, and designed to examine the effects of processing depth on recall. Through the execution of some database code, the data it generated were processed, aggregated, and then available to the students for analysis via a shared network folder. To enable evaluation, a parallel didactic version of the exercise was developed, in which a tutor described the experiment (with the aid of PowerPoint slides), and the students analysed a canned data set. Classes were then randomised to the two versions of the exercise, and students were invited to complete a post-workshop evaluation questionnaire in their own time. A series of generalized linear mixed models indicated that, compared to students in the didactic/canned condition,

students who participated in the experiment then analysed their own data displayed significantly greater knowledge of the methodological and statistical issues addressed in class, and were more confident regarding their ability to use this knowledge appropriately in the future. However, the two groups did not differ in terms of their subjective evaluation of, nor satisfaction with the workshop. These findings are consistent with interdisciplinary research indicating a positive association between the classroom implementation of active learning activities and student performance (Freeman et al., 2014), as well as literature suggesting that performance is not always clearly associated with satisfaction (Sizemore & Lewandowski, 2009).

Conducting Original Research

In addition to regular participation in class exercises like the one described by Allen and Baughman (2016), there is a general consensus that undergraduate psychology students ought to be regularly engaged in the full research process, from the development of meaningful research questions and hypotheses, through design, data collection, analysis, interpretation, and reporting. This consensus is evident in the learning goals, standards, and graduate attributes specified for undergraduate psychology degrees (APA 2016; APAC, 2010; BPS, 2015), and the frequency of ‘research project’ type assessments through the undergraduate curriculum (Kierniesky, 2005; Perlman & McCann, 2005; Stoloff et al., 2015). In lower years, students typically gain experience with the research process via class projects, which may address either novel research questions, or replicate well established psychological phenomena (Ball & Pelco, 2006; Bauer & Bennett, 2003; Chapdelaine & Chapman, 1999; Grahe et al., 2012; Holmes & Beins, 2015; Kim-Prieto & D’Oriano, 2011; Landrum & Smith, 2007; Larkin & Pines, 2005; Marek, Christopher, & Walker, 2004). Throughout the undergraduate degree, particularly in the United States, many students also participate in Undergraduate Research Experience (URE) programs or research assistantships (Craney, McKay, Mazzeo, Morris, Prigodich, & de Groot, 2011; Davis, 2007; Holmes & Beins, 2011; Kardash, 2000; Kierniesky, 2005; Landrum & Nelsen, 2002; Miller, 2015; Miller et al., 2008; Vespia, Wilson-Doenges, Martin, & Radosevich, 2012; Wayment & Dickson, 2008; Woodzicka, Ford, Caudill, & Ohanmamooreni, 2015). At the end of the undergraduate degree, the final, and most substantial research experience is the final year dissertation project, undertaken either individually or in small groups, and under the direct supervision of

a faculty member (APAC, 2010; BPS, 2015; Chew, 2015; F. Martin, Cranney, & Varcin, 2013; Roberts 2015a, 2015b). Lynne Roberts and I have supervised many such projects, and our reflections on the supervision process, the manner in which we've observed students approach their research projects, and the nature of the research we frequently engage them in (which is often reliant on the use of online questionnaires), have raised a number of issues that we have subsequently sought to investigate further.

The first of these relates to concerns about the relative ease with which online questionnaire data can be fabricated or falsified by research students eager to (a) reach the minimum sample sizes 'mandated' by their *a priori* power analyses; and (b) ultimately 'reject the null hypothesis'. Data fabrication refers to "making up data", whereas falsification involves "manipulating ... changing or omitting data" (Public Health Service Policies on Research Misconduct, 2005). A number of audits of published research and retraction notices (Claxton, 2005; Fang, Steen, & Casadevall, 2012; Grieneisen & Zhang, 2012; Madlock-Brown & Eichmann, 2015; Steen, 2011; Steneck, 2006; Wagner & Williams, 2011), as well as surveys of professional researchers regarding both their personal and colleagues' research (mis-)conduct (Fanelli, 2009; John, Loewenstein, & Prelec, 2012; Ranstam et al., 2000; Swazey, Anderson, & Lewis 1993; Titus, Wells, & Rhoades, 2008; Wells, 2008; Williams & Roberts, 2016) have suggested concerning rates of data fabrication and falsification. Falsified or fabricated data on the public record are problematic because they distort scientific knowledge and the decisions it informs (Steneck, 2006). They can also give the researchers involved an unfair competitive advantage, to the extent that their behaviour remains undiscovered (see Cyranoski, 2006; Gross, 2016; Stroebe, Postmes, & Spears, 2012). Research indicates that self-reported rates of data fabrication and falsification amongst students are considerably higher than those of professional researchers (Brimble & Stevenson-Clarke, 2005; Davidson, Cate, Lewis, & Hunter, 2000; Franklyn-Stokes & Newstead, 1995; Lawson, Lewis, & Birk, 1999/2000; McCabe, 2005, Rajah-Kanagasabai & Roberts, 2015; Swazey et al., 1993; Yang, 2012). Like professional researchers who are willing to engage in research misconduct, students who falsify or fabricate data also gain a competitive advantage relative to their peers. Furthermore, such behaviour is likely to result in impoverished learning (Brimble & Stevenson-Clarke, 2005), generalise to other contexts (Nonis & Swift, 2001; Stone, Jawahar, & Kisamore, 2011) and, when exposed, may undermine public

trust in science and higher education (Marsden, Carroll, & Neill, 2005). Relative to the available self-report data, objective data are considerably more limited, and only estimate the prevalence of student ‘cheating’ (though not data fabrication or falsification specifically) in a handful of esoteric contexts (e.g., Karlins, Michaels, & Podlogar, 1988; D. Martin, Rao, & Sloan, 2009; Pullen, Ortloff, Casey, & Payne, 2000; Ward & Beck, 1990). The absence of published studies that directly and objectively measure the prevalence of data fabrication or falsification in student populations is not unsurprising, considering the challenges involved in conducting such research. However, there are numerous indicators of *potential* fabrication and falsification that can be easily quantified. One such indicator is the presence of partially or completely duplicated (and thus identical) cases in a raw data set (Blasius & Thiessen, 2012). At Curtin University, psychology honours students are required to submit their raw data as digital appendices to their final dissertations. Consequently, the fourth aim of the current thesis was:

To (a) systematically analyse a population of psychology honours students’ final, submitted data sets for evidence of data duplication; and (b) where such duplication is found, systematically analyse the relevant students’ dissertations for indicators of its likely aetiology.

Honours is the primary pathway to both specialised professional (e.g., clinical, counselling, organisational psychology) and research careers in psychology. The dissertation project is the first major research experience for most honours students, and undesirable behaviours developed during this formative experience may generalise. At the time the research by Allen, Lourenco, and Roberts (2016, paper 4) was conducted, we were not aware of any research attempting to objectively measure undergraduate student research practices suggestive of data fabrication or falsification. Consequently, we used techniques developed by Blasius and Thiessen (2012) to systematically examine 18 psychology honours students’ final, submitted dissertation data sets for evidence of data duplication, which is one indicator of possible data fabrication. When such evidence was detected, we examined the students’ submitted dissertations for indicators of its likely causes. Although we didn’t identify any completely duplicated cases, there were numerous partial duplicates. Rather than indicating fabrication, however, these partial duplicates were more likely a consequence of poor measure selection, insufficient data screening, and/or a range of

participant characteristics. These findings prompted several suggestions regarding the teaching and supervision of undergraduate student researchers, including (a) emphasising the importance of selecting measures valid for the populations with which they are being used (and pre-testing in times of uncertainty); (b) providing practical tutelage on the data screening/cleaning process; (c) talking regularly with students about research integrity (which Schoenherr, 2015, argues, currently receives variable and inconsistent coverage across psychology courses); and (d) de-emphasising the importance ' $p < .05$ '.

Allen and Roberts (2016; paper 5) was also a consequence of pragmatic concerns emerging during the supervision of students' final-year research projects. At the time this research was undertaken, online surveying software was relatively undeveloped (with the market dominated by SurveyMonkey.com; see Allen & Roberts, 2010), and securing official permission to use Curtin University branding in undergraduate student research was a reasonably cumbersome process. This inevitably led to the question, 'is it worth it?' In other words, would securing branding rights (including the right to host information about the study on official University websites), then finding and purchasing access to online surveying software which could accommodate prominent branding (a feature that was not available in SurveyMonkey.com free accounts at the time), have a measurable impact on data quality? The indices of data quality that Allen and Roberts (2016) focused on were survey drop-out (which occurs when a respondent commences, but does not complete a survey; Lozar Manfreda & Vehovar, 2002) and item non-response (when a respondent skips a question they have been exposed to, and are otherwise eligible to complete; Bosnjak & Tuten, 2001), leading to the fifth aim of the current thesis:

To systematically manipulate the prominence of university sponsorship on a web survey, then measure the effects of this manipulation on survey drop-out and item non-response.

Web survey drop-out and item non-response are known to be influenced by, and associated with a number of survey and respondent related characteristics (e.g., Ekman, Klint, Dickman, Adami, & Litton, 2007; Goritz, 2010; Kays, Gathercoal, & Buhrow, 2012; B. Messer, Edwards, & Dillman, 2012; Nosek, Sriram, & Umansky, 2012; O'Neil, Penrod, & Bornstein, 2003; Sanchez-Fernandez, Munoz-Leiva, & Montoro-Rios, 2012; Stieger, Reips, & Voracek, 2007; T. Yan, Conrad, Tourangeau,

& Couper, 2011). Survey sponsorship (i.e., the prominence and/or nature of corporate/university branding on a survey instrument; Boulianne, 2008) is also known to influence survey completion behaviour in offline contexts (Edwards et al., 2002; Fox, Crask, & Kim, 1988; Peterson, 1975). However, when the research reported in Allen and Roberts (2016) was undertaken, little was known about effects of survey sponsorship on drop-out and item non-response online, and the two studies which had investigated these relationships produced inconsistent findings (Boulianne, Klofstad, & Basson, 2011; Heerwegh & Loosveldt, 2006).

In the first of two studies, Allen and Roberts (2016) randomised 498 participants to online surveys with either high or low university sponsorship. There was no difference between the proportions of participants dropping out of each condition. However, counter to our predictions, participants in the high sponsorship condition displayed significantly higher item non-response. In Study 2 ($N = 159$), which addressed a rival explanation for the findings in Study 1, sponsorship prominence had no impact on either outcome variable. Overall, we argued that these findings suggest that hosting information pages on university websites, placing university logos on survey pages, and including the name of the university in survey URLs (Uniform Resource Locators, or world-wide-web addresses) do not reliably impact on drop-out or item non-response. As supervisors of undergraduate research students without ready access to university web servers or branding (at the time), these findings gave us comfort, as they indicated that minimally visible sponsorship does not necessarily compromise data quality.

Allen and Roberts (2010; paper 6) and Roberts and Allen (2015; paper 7) are also tightly linked to the use of online surveys in research involving students (as researchers, participants, or both). The first, Allen and Roberts (2010), is a discussion paper around the ethics of outsourcing online survey research; a practice which appears at least as common now as it was when the paper was written (see <https://www.socialpsychology.org/expts.htm>, <http://psych.hanover.edu/research/exponnet.html> etc.). In 2009/10, the ethical issues associated with outsourcing the design and/or hosting of online surveys to external, for-profit service providers (e.g., Qualtrics.com, SurveyMonkey.com etc.) were relatively unexplored. Yet we, and colleagues across the higher education sector (Beiderniki & Kerschbaumer, 2007; Buchanan & Hvizdak, 2009; Gaiser & Schreiner,

2009; Kaczmirek, 2008; Sue & Ritter, 2007; Wright, 2005), were using and recommending these providers to both students and peers with considerable regularity. Hence, the aim of Allen and Roberts (2010), as well as the sixth aim of the current thesis was:

To review the key ethical concerns associated with the outsourcing of web survey design and hosting, and offer best practice guidelines regarding the use of for-profit web survey providers for the purposes of collecting research data.

The specific ethical issues raised by Allen and Roberts (2010) included (a) the potential consequences to participant confidentiality and/or anonymity associated with providers' data protection and transmission practices, routine collection of potentially personally identifying information from survey respondents, and privacy and legal disclosure policies; and (b) concerns surrounding the potential for outsourcing to undermine the credibility of academic research and quality of research data. The best practice recommendations we offered included (a) encouraging research institutions to thoughtfully develop institutional web surveying policies and procedures, and to endorse a specific policy-compliant 'preferred' web-surveying provider; (b) 'sandwiching' outsourced surveys between information sheets and debriefing materials hosted on institutional websites; and (c) separating the collection of personal identifiers (e.g., for consent or competition entry purposes) from the collection of research data, and using locally hosted secure web-forms for the former, wherever possible.

Our second contribution to the dialogue around the ethics of online surveying was Roberts and Allen (2015), in which we narrowed our focus to the ethical use of online surveys in educational research undertaken specifically in higher education contexts. The aim of this paper, and the seventh aim of the current thesis was:

To outline and illustrate the key ethical issues associated with the use of web surveys in the conduct of educational research in higher education settings, and to offer ethically defensible recommendations for the use of web surveys in this context.

The ethical issues we described, illustrated (with reference to our own prior research), and provided practical recommendations for addressing in Roberts and Allen (2015) included: (a) navigating the dual roles that arise when seeking to conduct

online survey research with one's own students and/or colleagues as participants; (b) ensuring that the consent provided by participants is both informed and voluntary; (c) the use of incentives; (d) maintaining participants' privacy, confidentiality and/or anonymity; and (e) maximising the likelihood of obtaining quality data. We further advocated the adoption of a situated/process approach to ethical thinking and action (Guillemin & Gillam, 2004; James & Busher, 2007; Simons & Usher, 2000), and ultimately argued that online surveying, when deployed appropriately, offers an ethically defensible method for conducting educational research that would simply not be feasible offline.

Challenges in Conducting Research

Regardless of their level of research experience, many undergraduate psychology students find quantitative data analysis to be one of the most challenging aspects of conducting research (Murtonen et al., 2008). Furthermore, research suggests that students particularly struggle with the development of "selection skills" (Ware & Chastain, 1989, p. 222), or the identification of appropriate statistical tests and procedures for different types of research questions, hypotheses, and data types (e.g., Gardner & Hudson, 1999). However, these skills are trainable (Ware & Chastain, 1991; J. Yan & Lavigne, 2014), and appear underpinned, at least partially, by "structural awareness" (Quilici & Mayer, 2002, p. 326), which reflects an ability to ignore the surface features of a research problem, and instead concentrate on its structural features and the relationships between them. Whist even relatively experienced students (e.g., at Honours, Masters, and PhD level) find this process challenging (e.g., Gardner & Hudson, 1999; Rabinowitz & Hogan, 2008), "experts" do not. Beyond the focus on surface and structural components of research scenarios, little is known about how students and experts select statistics. Hence the eighth aim of the current thesis was:

To develop a rich account of the strategies that psychology students and academics adopt when selecting statistical tests or procedures appropriate to different research questions, hypotheses and data types.

Knowing that a range of aids and resources have been developed to facilitate the task of selecting statistical tests and procedures (e.g., Beitz, 1998; Carlson, Protsman, & Tomaka, 2005; Koch & Gobell, 1999; Protsman & Carlson, 2008; Twycross & Shields, 2004), that pedagogic resources are more likely to be promoted

by instructors and adopted by students if they are developed with the expressed needs and preferences of these stakeholder groups in mind, and that these needs and preferences are currently not well understood, the ninth aim of the current thesis was:

To elicit students' and academics' views on the nature of resources that could facilitate the statistical decision making process.

Aims eight and nine were achieved in Allen, Dorozenko, and Roberts (2016; paper 8). The research described in this paper was undertaken in two phases. In the first phase, nine psychology undergraduates (all of whom had completed at least one quantitative methods unit) were shown a series of brief research scenarios, and asked to describe the process they would follow to identify an appropriate statistical test or procedure for each. Thematic analysis indicated that all found this task difficult, and even those who had completed several research methods units struggled to articulate how they would approach the scenarios on more than a very superficial level. Although some of the students recognised the existence of a systematic decision making process that can be followed when selecting statistical tests and procedures, none could describe it clearly or completely.

In the second phase of the research, we presented the same scenarios to 10 psychology academics, who each had particular expertise in conducting research and/or research methods instruction. Predictably, these 'experts' 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 pays consideration to numerous contextual factors. The academics were sensitive to the challenges students experience when making statistical decisions, which they partially attributed to the manner in which research methods and statistics are commonly taught. When we asked both groups to consider the format and features of an aid that could facilitate the statistical decision making process, they generally expressed a preference for an accessible, comprehensive, and reputable resource, which follows a basic decision tree logic. From the perspective of the academics in particular, this aid should function primarily as a teaching tool, which engages the user with each choice-point in the decision making process, rather than merely pointing to an "answer".

The findings presented in Allen, Dorozenko, and Roberts (2016) helped inform the development of StatHand, a free, cross-platform application designed to support

students' statistical decision making. The final aim of the current thesis, achieved with Allen et al. (2016; paper 9), was:

To (a) describe the theoretical and empirical rationales behind the development of StatHand; (b) outline the feature set of the application; and (c) articulate a series of evidence-based guidelines for integrating the use of StatHand into the research methods curriculum.

The StatHand application guides users through a series of simple, annotated questions to help them identify a statistical test or procedure appropriate to their circumstances. In taking this approach, it prompts the user to sequentially consider each structural component of their research relevant to identifying an appropriate analytic strategy. Thus, StatHand is functionally comparable to a paper-based decision tree, a type of graphic organiser that has long been popular amongst both statistics educators and students (Allen, Bennett, & Heritage, 2014; Carlson et al., 2005; Fok, Angelidis, Ibrahim, & Fok, 1995; Protsman & Carlson, 2008; Tabachnick & Fidell, 2013). Such decision trees make explicit the interconnectedness (and differentiation) amongst statistical concepts, and provide an organisational framework for their integration into, and subsequent retrieval from long-term memory (Schau & Mattern, 1997; Yin, 2012). Research has demonstrated the efficacy of paper-based decision trees in terms of both statistical decision making speed and accuracy (Carlson et al., 2005; Protsman & Carlson, 2008). However, despite their popularity and utility, paper-based decision trees have limitations, primarily associated with the need to fit them onto a single sheet of paper, or page in a textbook. Hypertext trees overcome many of these limitations, but require a constant internet connection. Conversely, mobile applications like StatHand can maintain all/most functionality without an internet connection (Kretser et al., 2015). Furthermore, smart device penetration is near ubiquitous amongst higher education students (Chen, Sellhamer, Bennett, & Bauer, 2015; Dahlstrom & Bichsel, 2014), who express a preference for using these devices for various educational purposes (Bowen & Pistilli, 2012; Farley et al., 2015).

StatHand is currently available in the iOS app store for both iPhone and iPad. Furthermore, a web-application replicating its feature set and formatted for all other devices (e.g., Android devices, desktop computers etc.) can be found at <https://stathand.net/>. Our recommendations for the integration of StatHand into research methods curricula were informed by the Unified Theory of Acceptance and

Use of Technology (Venkatesh, Morris, Davis, & Davis, 2003). They include (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 both methods and non-methods classes.

Summary of Thesis Objective and Aims

The overarching objective of the current thesis was *to probe issues and strategies linked to the engagement of undergraduate psychology students with research methods, and the process of conducting research*. This objective is manifest in 10 specific aims, which are summarised in Table 1. Table 1 also identifies the papers in which these aims are addressed and met.

Table 1

Thesis Aims and the Papers in Which They Were Addressed

Aim	Paper
1 To develop a rich, qualitative account of undergraduate psychology students' perspectives on the educational value of research participation.	1 Roberts, L. D., & Allen, P. J. (2012). Student perspectives on the value of research participation. In S. McCarthy, K. L. Dickson, J. Cranney, A. Trapp, & V. Karandashev (Eds.), <i>Teaching psychology around the world</i> (vol 3, pp. 198-211). Newcastle, England: Cambridge Scholars.
2 To develop a brief, reliable and valid measure of the perceived educational value of research participation, grounded in qualitative perspectives of student research participants, as captured by Roberts and Allen (2012).	2 Roberts, L. D., & Allen, P. J. (2013). A brief measure of student perceptions of the educational value of research participation. <i>Australian Journal of Psychology</i> , 65, 22-29. doi:10.1111/ajpy.12007
3 To (a) develop a class exercise in which students participate in an experiment, engage in class discussion around its methods, then use class-generated data to practice various data handling and statistical procedures; and (b) evaluate the aforementioned exercise in terms of (i) its subjective appeal to students; and (ii) its pedagogic effectiveness.	3 Allen, P. J., & Baughman, F. D. (2016). Active learning in research methods classes is associated with higher knowledge and confidence, though not evaluations or satisfaction. <i>Frontiers in Psychology</i> , 7, Article 279. doi:10.3389/fpsyg.2016.00279

-
- | | | | |
|---|---|---|---|
| 4 | To (a) systematically analyse a population of psychology honours students' final, submitted data sets for evidence of data duplication; and (b) where such evidence is found, systematically analyse the relevant students' dissertations for indicators of its likely aetiology. | 4 | <u>Allen, P. J.</u> , Lourenco, A., & Roberts, L. D. (2016). Detecting duplication in students' research data: A method and illustration. <i>Ethics & Behavior</i> , 26, 300-311. doi:10.1080/10508422.2015.1019070 |
| 5 | To systematically manipulate the prominence of university sponsorship on a web survey, then measure the effects of this manipulation on survey drop-out and item non-response. | 5 | <u>Allen, P. J.</u> , & Roberts, L.D. (2016). The impact of academic sponsorship on Web survey dropout and item non-response. <i>First Monday</i> , 21(2). doi:10.5210/fm.v21i2.6144 |
| 6 | To review the key ethical concerns associated with the outsourcing of web survey design and hosting, and offer best practice guidelines regarding the use of for-profit web survey providers for the purposes of collecting research data. | 6 | <u>Allen, P. J.</u> , & Roberts, L. D. (2010). The ethics of outsourcing online survey research. <i>International Journal of Technoethics</i> , 1, 35-48. doi:10.4018/jte.2010070104 |
| 7 | To outline and illustrate the key ethical issues associated with the use of web surveys in the conduct of educational research in higher education settings, and to offer ethically defensible recommendations for the use of web surveys in this context. | 7 | Roberts, L. D., & <u>Allen, P. J.</u> (2015). Exploring ethical issues associated with using online surveys in educational research. <i>Educational Research and Evaluation</i> , 21, 95-108. doi:10.1080/13803611.2015.1024421 |
-

-
- 8 To develop a rich account of the strategies that psychology students and academics adopt when selecting statistical tests or procedures appropriate to different research questions, hypotheses and data types. 8 **Allen, P. J.**, Dorozenko, K. P., & Roberts, L. D. (2016) Difficult decisions: A qualitative exploration of the statistical decision making process from the perspectives of psychology students and academics. *Frontiers in Psychology, 7*, Article 188. doi:10.3389/fpsyg.2016.00188
- 9 To elicit students' and academics' views on the nature of resources that could facilitate the statistical decision making process.
- 10 To (a) describe the theoretical and empirical rationales behind the development of StatHand; (b) outline the feature set of the application; and (c) articulate a series of evidence-based guidelines for integrating the use of StatHand into the research methods curriculum. 9 **Allen, P. J.**, Roberts, L. D., Baughman, F. D., Loxton, N. J., Van Rooy, D., Rock, A. J., & Finlay, J. (2016). Introducing StatHand: A cross-platform mobile application to support students' statistical decision making. *Frontiers in Psychology, 7*, Article 288. doi:10.3389/fpsyg.2016.00288
-

Conclusion

Combined with this exegesis, the nine papers in this thesis offer a range of insights into, and strategies that promote the engagement of undergraduate psychology students with research methods, and with the process of conducting research. The work contained herein is consciously and deliberately ‘applied’, was born from reflective practice, and has informed pedagogic practice at local and national levels.

The first two papers (Roberts & Allen, 2012, 2013) enhance our understanding of undergraduate psychology students’ perspectives on the educational value of participating in research, and provide a psychometrically sound measure of those perceptions. Undergraduate students have long been our (i.e., the discipline of psychology’s) primary sampling frame (Arnett, 2008; Higbee et al., 1982; Korn, 1999; McNemar, 1946; Wintre et al., 2001). Despite ongoing concerns around external validity (Henrich, Heine, & Norenzayan, 2010; Sears, 1986), this seems unlikely to change anytime soon (e.g., Henry, 2008). However, the regulatory environment in which we sample from this population is changing. Universities increasingly see students as ‘clients’ or ‘customers’, and customer satisfaction increasingly drives administrative decision-making across the higher education sector (McGhee, 2015; Sharrock, 2013). There is also an increasing awareness of, and sensitivity to the ethics associated with ‘students as participants’ (Dalziel, 1996; Sieber, 1999; Tabachnick et al., 1991). Combined, these factors compel researchers to acknowledge the contribution that student-participants make to the progress of psychological science, and to reciprocate with the provision of meaningful learning opportunities. Roberts and Allen (2012) suggested strategies toward achieving this goal, whilst Roberts and Allen (2013) offered a method for quantifying their efficacy. This work also informed the proposal to incorporate a participant pool scheme into the second year of the undergraduate psychology program at Curtin in 2012. This scheme has subsequently been expanded into the first and third years of the program, and currently generates over 3000 hours of participation experience (for students) and data (for both student and faculty researchers) annually.

Papers 3-7 (Allen & Baughman, 2016; Allen, Lourenco, & Roberts, 2016; Allen & Roberts, 2010, 2016; Roberts & Allen, 2015) have similarly achieved local impact. As a consequence of the scholarly inquiry behind Allen and Baughman (2016), students in the undergraduate psychology program at Curtin now have expanded

opportunities to work with authentic, personally-meaningful data throughout their course. Allen, Lourenco, and Roberts (2016) highlighted several gaps in the year four research methods curriculum, which have subsequently been filled. Allen and Roberts (2010, 2016) and Roberts and Allen (2015) helped establish local standards and protocols for the conduct of online research. Finally, Allen, Dorozenko, and Roberts (2016; paper 8) directly informed the development of StatHand (Allen et al., 2016; paper 9), which is currently embedded in the psychology research methods curricula at several Australian universities. Work to establish the efficacy of StatHand is ongoing, but preliminary results suggest that the application promotes greater decision making accuracy, and is instructionally efficient, relative to more traditional statistical decision making aids (Allen et al., 2015).

Despite their impact and novelty, and the extent to which they contribute to the evidence-base supporting a range of teaching practices, the papers contained herein are not without limitations. Many of these parallel limitations common across the psychology SoTL (Scholarship of Teaching and Learning) literature (see Gurung, 2015; Wilson-Doenges & Gurung, 2013; Wilson-Doenges, Troisi, & Bartsch, 2016). SoTL in psychology has been defined by Gurung, Ansborg, Alexander, Lawrence, and Johnson (2008) as:

Literature-based inquiry into processes and outcomes involved in the teaching and learning of psychology. When appropriate, the activity must follow the standards and practices delineated by the scientific method (e.g., systematic observations, well-developed operational definitions, accurate statistical analyses). The activity generates a product that is peer-reviewed on the basis of whether that product contributes new knowledge to the field and/or invites conceptual replication and must yield a publicly presented product.

For example, the empirical work presented in this thesis is cross-sectional, rather than longitudinal (e.g., Allen & Baughman, 2016; Allen, Dorozenko, & Roberts, 2016; Allen, Lourenco, & Roberts, 2016; Allen & Roberts, 2016; Roberts & Allen, 2012, 2013), and thus has little to say about change or development over time. Ideally, the participants in Allen and Baughman (2016) would have been randomised to treatment conditions, allowing for an experimental test of each hypothesis. However, randomisation is not possible in such a field study in a contemporary tertiary environment, where students self-select into tutorial classes based on a range of

personal considerations. Consequently, our design was quasi-experimental (a comparison of intact groups, randomised to treatment conditions), and our capacity for causal inference was limited. Furthermore, the samples in several of the studies described herein were relatively small (Allen & Baughman, 2016; Allen, Lourenco, & Roberts, 2016; Roberts & Allen, 2013) and most were sourced from a single higher education institution (and often a single course; Allen & Baughman, 2016; Allen, Lourenco, & Roberts, 2016; Roberts & Allen, 2012, 2013). A reliance on small, homogenous samples carries implications for both statistical power and external validity. Finally, although a variety of methods were used across the six empirical papers in this thesis, each makes use of just one. The use of longitudinal designs, experimentation, large and diverse samples, and mixed methods are five of the eight ‘gold standard benchmarks’ to which Wilson-Doenges and colleagues (2016) argue all SoTL practitioners in psychology should aspire.

The remaining gold standards proposed by Wilson-Doenges and colleagues (2016) are (a) situating all SoTL in a theoretical and/or empirical context; (b) making use of advanced statistical methods; and (c) maintaining high ethical standards. Throughout this thesis there are multiple illustrations of each. For example, the work in Allen, Dorozenko, and Roberts (2016) and Allen et al. (2016) was informed by empirical work demonstrating the paucity of ‘selection skills’ amongst many students (Gardner & Hudson, 1999; Ware & Chastain, 1989, 1991), the theoretical construct of ‘structural awareness’ (Quilici & Mayer, 2002), and previous work demonstrating the efficacy of decision trees (Carlson et al., 2005; Protsman & Carlson, 2008). Generalized linear mixed modelling was used in Allen and Baughman (2016) to account for non-independence of observations, a problem routinely encountered (and less commonly addressed) when analysing data collected across multiple classes, teachers, courses and/or schools (Murray & Hannan, 1990). Finally, the recommendations provided in Roberts and Allen (2015) for navigating multiple ethical issues associated with conducting educational research were routinely followed, where applicable, in the conduct of the empirical work documented in this thesis.

Although the directions for future research stemming from the individual studies presented in this thesis are diverse, they are united by a common theme. That is, the need to continue the development of a high quality evidence-base underpinning

teaching methods and techniques that promote undergraduate psychology students' deep engagement with research methods, and with the process of conducting research.

References

- Addison, W. E., Stowell, J. R., & Reab, M. D. (2015). Attributes of introductory psychology and statistics teachers: Findings from comments on RateMyProfessors.com. *Scholarship of Teaching and Learning in Psychology, 1*, 229-234. doi:10.1037/stl0000034
- Allen, P., Bennett, K., & Heritage, B. (2014). *SPSS Statistics version 22: A practical guide*. Melbourne, Australia: Cengage Learning.
- Allen, P. J., & Baughman, F. D. (2016). Active learning in research methods classes is associated with higher knowledge and confidence, though not evaluations or satisfaction. *Frontiers in Psychology, 7*, Article 279. doi:10.3389/fpsyg.2016.00279
- Allen, P. J., Dorozenko, K. P., & Roberts, L. D. (2016) Difficult decisions: A qualitative exploration of the statistical decision making process from the perspectives of psychology students and academics. *Frontiers in Psychology, 7*, Article 188. doi:10.3389/fpsyg.2016.00188
- Allen, P. J., Lourenco, A., & Roberts, L. D. (2016). Detecting duplication in students' research data: A method and illustration. *Ethics & Behavior, 26*, 300-311. doi:10.1080/10508422.2015.1019070
- Allen, P. J., & Roberts, L. D. (2010). The ethics of outsourcing online survey research. *International Journal of Technoethics, 1*, 35-48. doi:10.4018/jte.2010070104
- Allen, P. J., & Roberts, L.D. (2016). The impact of academic sponsorship on Web survey dropout and item non-response. *First Monday, 21*(2). doi:10.5210/fm.v21i2.6144
- Allen, P. J., Roberts, L. D., Baughman, F. D., Loxton, N. J., Van Rooy, D., Rock, A. J., & Finlay, J. (2015, October). *StatHand: A mobile application to support students' statistical decision making*. Poster presented at the 2015 International Society for the Scholarship of Teaching and Learning (ISSoTL) conference, Melbourne, Australia.
- Allen, P. J., Roberts, L. D., Baughman, F. D., Loxton, N. J., Van Rooy, D., Rock, A. J., & Finlay, J. (2016). Introducing StatHand: A cross-platform mobile

- application to support students' statistical decision making. *Frontiers in Psychology*, 7, Article 288. doi:10.3389/fpsyg.2016.00288
- American Psychological Association. (2016). Guidelines for the undergraduate psychology major: Version 2.0. *American Psychologist*, 71, 102-111. doi:10.1037/a0037562
- American Psychological Association Board of Educational Affairs Task Force on Psychology Major Competencies. (2013). *APA guidelines for the undergraduate psychology major: Version 2.0*. Washington, DC: Author. Retrieved from <http://www.apa.org/ed/precollege/about/psymajor-guidelines.pdf>
- American Psychological Association Presidential Task Force on Evidence Based Practice. (2006). Evidence-based practice in psychology. *American Psychologist*, 61, 271–285. doi:10.1037/0003-066X.61.4.271
- Arnett, J. J. (2008). The neglected 95%: Why American psychology needs to become less American. *American Psychologist*, 63, 602-614. doi:10.1037/0003-066X.63.7.602
- Australian Psychology Accreditation Council. (2010). *Rules for accreditation and accreditation standards for psychology courses: Version 10*. Melbourne, Australia, Author. Retrieved from https://www.psychologycouncil.org.au/Assets/Files/APAC_Rules_for%20Accreditation_and_Accreditation_Standards_for%20Psychology_Courses_Ver_10_June_2010.pdf
- Australian Psychology Accreditation Council. (2016). *Proposed accreditation standards for psychology programs: Consultation paper June 2016*. Melbourne, Australia: Author. Retrieved from http://admin.psychologycouncil.org.au/Assets/Files/APAC_Consultation-Paper-020616.pdf
- Ball, C. T., & Pelco, L. E. (2006). Teaching research methods to undergraduate psychology students using an active cooperative learning approach. *International Journal of Teaching and Learning in Higher Education*, 17, 147-154. Retrieved from <http://www.isetl.org/ijtlhe/pdf/IJTLHE38.pdf>

- Bartholomay, E. M., & Sifers, S. K. (2016). Student perception of pressure in faculty-led research. *Learning and Individual Differences*. Advance online publication. doi:10.1016/j.lindif.2016.08.025
- Bauer, K. W., & Bennett, J. S. (2003). Alumni perceptions used to assess undergraduate research experience. *Journal of Higher Education*, 74, 210-230. doi:10.1353/jhe.2003.0011
- Beiderniki, G., & Kerschbaumer, A. (2007). Comparison of online survey tools. In R. A. Reynolds, R. Woods, & J. D. Baker (Eds.), *Handbook of research on electronic surveys and measurements* (pp. 264–272). Hershey, PA: IGI Global.
- Beitz, J. M. (1998). Helping students learn and apply statistical analysis: A metacognitive approach. *Nurse Educator*, 23, 49–51. doi:10.1097/00006223-199801000-00016
- Benjamin, L. T. Jr. (2014). *A brief history of modern psychology* (2nd ed.). Hoboken, NJ: Wiley Blackwell.
- Blasius, J., & Thiessen, V. (2012). *Assessing the quality of research data*. London, UK: Sage.
- Bosnjak, M., & Tuten, T. L. (2001). Classifying response behaviors in web-based surveys. *Journal of Computer Mediated Communication*, 6(3). doi:10.1111/j.1083-6101.2001.tb00124.x
- Boulianne, S. (2008). Survey sponsor. In P. J. Lavrakas (Ed.), *Encyclopedia of survey research methods* (pp. 868-869). Thousand Oaks, CA: Sage.
- Boulianne, S., Klofstad, C. A., & Basson, D. (2011). Sponsor prominence and response patterns to an online survey. *International Journal of Public Opinion Research*, 23, 79-87. doi:10.1093/ijpor/edq026
- Bowen, K., & Pistilli, M. D. (2012). *Student preferences for mobile app usage* (Research Bulletin). Louisville, CO: Educause Center for Analysis and Research. Retrieved from <https://net.educause.edu/ir/library/pdf/ERB1210.pdf>
- Brimble, M., & Stevenson-Clarke, P. (2005). Perceptions of the prevalence and seriousness of academic dishonesty in Australian universities. *The Australian Educational Researcher*, 32(3), 19–44. doi:10.1007/BF03216825

- British Psychological Society (2015). *Standards for the accreditation of undergraduate, conversion and integrated masters programmes in psychology (October 2015)*. Leicester, England: British Psychological Society. Retrieved from http://www.bps.org.uk/system/files/Public%20files/PaCT/undergraduate_accr_education_2015_web.pdf
- Brysbaert, M., & Rastle, K. (2013). *Historical and conceptual issues in psychology* (2nd ed.). Harlow, England: Pearson.
- Brody, J. L., Gluck, J. P., & Aragon, A. S. (2000). Participants' understanding of the process of psychological research: Debriefing. *Ethics & Behavior, 10*, 13-25. doi:10.1207/S15327019EB1001_2
- Buchanan, E. A., & Hvizdak, E. E. (2009). Online survey tools: Ethical and methodological concerns of Human Research Ethics Committees. *Journal of Empirical Research on Human Research Ethics, 4*, 37-48. doi:10.1525/jer.2009.4.2.37
- Carlson, M., Protsman, L., & Tomaka, J. (2005). Graphic organizers can facilitate selection of statistical tests. Part 1 - analysis of group differences. *Journal of Physical Education Therapy, 19*, 57-66. Retrieved from <http://aptaeducation.org/members/jopte/index.cfm>
- Chapdelaine, A., & Chapman, B. L. (1999). Using community-based research projects to teach research methods. *Teaching of Psychology, 26*, 101-105. doi:10.1207/s15328023top2602_4
- Chen, B., Sellhamer, R., Bennett, L., & Bauer, S. (2015). Students' mobile learning practices in higher education: A multi-year study. *Educause Review*. Retrieved from <http://er.educause.edu/articles/2015/6/students-mobile-learning-practices-in-higher-education-a-multiyear-study>
- Chew, S. L. (2015). Directing undergraduate research in independent studies, honors, and thesis projects. In D. S. Dunn (Ed.), *The Oxford handbook of undergraduate psychology education* (pp. 241-253). New York: Oxford University Press.

- Ciarocco, N. J., Lewandowski, G. W. Jr., & Van Volkom, M. (2013). The impact of a multifaceted approach to teaching research methods on students' attitudes. *Teaching of Psychology, 40*, 20-25. doi:10.1177/0098628312465859
- Claxton, L. D. (2005). Scientific authorship part 1. A window into scientific fraud. *Mutation Research, 589*, 17–30. doi:10.1016/j.mrrev.2004.07.003
- Connors, F. A., McCown, S. M., & Roskos-Ewoldsen, B. (1998). Unique challenges in teaching undergraduate statistics. *Teaching of Psychology, 25*, 40-42. doi:10.1207/s15328023top2501_12
- Coulson, M. (1999). In at the deep end or still testing the water? Subject pools and required participation in UK psychology departments. *Psychology Teaching Review, 8*(1),19-25. Retrieved from <http://www.bps.org.uk/networks-and-communities/member-microsite/division-academics-researchers-teachers-psychology>
- Craney, C., McKay, T., Mazzeo, A., Morris, J., Prigodich, C., & de Groot, R. (2011). Cross-discipline perceptions of the undergraduate research experience. *The Journal of Higher Education, 82*, 92-113. doi:10.1353/jhe.2011.0000
- Cranney, J., & Dunn, D. S. (2011). Psychological literacy and the psychologically literate citizen: New frontiers for a global discipline. In J. Cranney & D. S. Dunn (Eds.), *The psychologically literate citizen: Foundations and global perspectives* (pp. 3-12). Oxford, England: Oxford University Press.
- Cranney, J., Morris, S., & Botwood, L. (2015). Psychological literacy in undergraduate psychology education. In D. S. Dunn (Ed.), *The Oxford handbook of undergraduate psychology education* (pp. 863-872). New York: Oxford University Press.
- Cranney, J., Turnbull, C., Provost, S. C., Martin, F., Katsikitis, M., White, F. A., ... Varcin, K. J. (2009). Graduate attributes of the 4-year Australian undergraduate psychology program. *Australian Psychologist, 44*, 253-262. doi:10.1080/00050060903037268
- Cromer, L. D., Reynolds, S. M., & Johnson, M. D. (2013). Who says: “no fair”? What personality and an experiment in educational value tell us about perceptions of costs and benefits of research pool requirements. *Journal of the Scholarship of*

- Teaching and Learning*, 13, 134-155. Retrieved from <http://josotl.indiana.edu/article/view/2037>
- Cyranoski, D. (2006). Verdict: Hwang's human stem cells were all fakes. *Nature*, 439, 122–123. doi:10.1038/439122a
- Dahlstrom, E., & Bichsel, J. (2014). *ECAR study of undergraduate students and information technology, 2014*. Louisville, CO: Educause Center for Analysis and Research. Retrieved from <https://net.educause.edu/ir/library/pdf/ss14/ERS1406.pdf>
- Dalziel, J. R. (1996). Students as research subjects: Ethical and educational issues. *Australian Psychologist*, 31, 119-123. doi:10.1080/00050069608260190
- Davidson, E. W., Cate, H. E., Lewis, C. M. Jr., & Hunter, M. (2000, November). *Data manipulation in the undergraduate laboratory: What are we teaching?* Paper presented at the First ORI Research Conference of Research Integrity, Bethesda, MD. Retrieved from <http://ori.dhhs.gov/documents/papers/davidson.pdf>
- Davis, J. R., & Fernald, P. S. (1975). Laboratory experience versus subject pool. *American Psychologist*, 30, 523-524. doi:10.1037/h0076758
- Davis, S. F. (2007). Student-faculty research collaboration: A model for success in psychology. In D. S. Dunn, R. A. Smith, & B. C. Beins (Eds.), *Best practices for teaching statistics and research methods in the behavioral sciences* (pp. 15-23). Mahwah, NJ: Lawrence Erlbaum Associates.
- Diamond, M. R., & Reidpath, D. D. (1992). Psychology ethics down under: A survey of student subject pools in Australia. *Ethics & Behavior*, 2, 101-108. doi:10.1207/s15327019eb0202_3
- Dunn, D. S. (2010). Teaching by doing: Making research methods active and engaging. In S. A. Meyers & J. R. Stowell (Eds.), *Essays from e-xcellence in teaching* (Vol. 9, pp. 11-16). Retrieved from <http://teachpsych.org/resources/e-books/eit2009/index.php>
- Dunn, D. S. (2015). The promise of psychology education: Some modest marching orders. In D. S. Dunn (Ed.), *The Oxford handbook of undergraduate psychology education* (pp. 883-886). New York: Oxford University Press.

- Dunn, D. S., Brewer, C. L., Cautin, R. L., Gurung, R. A. R., Keith, K. D., McGregor, L. N., ... Voigt, M. J. (2010). The undergraduate psychology curriculum: Call for a core. In D. F. Halpern (Ed.), *Undergraduate education in psychology: A blueprint for the future of the discipline* (pp. 47-61). Washington, DC: American Psychological Association. doi:10.1037/12063-003
- Dunn, P. K., Carey, M. D., Richardson, A. M., & McDonald, C. (2016). Learning the language of statistics: Challenges and teaching approaches. *Statistics Education Research Journal*, 15(1), 8-27. Retrieved from [http://iase-web.org/documents/SERJ/SERJ15\(1\)_Dunn.pdf](http://iase-web.org/documents/SERJ/SERJ15(1)_Dunn.pdf)
- Earley, M. A. (2014). A synthesis of the literature on research methods education. *Teaching in Higher Education*, 19, 242–253. doi:10.1080/13562517.2013.860105
- Edwards, P., Roberts, I., Clarke, M., DiGuseppi, C., Pratap, S., Wentz, R., & Kwan, I. (2002). Increasing response rates to postal questionnaires: Systematic review. *British Medical Journal*, 324, 1183-1191. doi:10.1136/bmj.324.7347.1183
- Ekman, A., Klint, A., Dickman, P. W., Adami, H. –O., & Litton, J. –E. (2007). Optimizing the design of web-based questionnaires – Experience from a population-based study among 50,000 women. *European Journal of Epidemiology*, 22, 293-300. doi:10.1007/s10654-006-9091-0
- Fanelli, D. (2009). How many scientists fabricate and falsify research? A systematic review and meta-analysis of survey data. *PLoS One*, 4, e5738. doi:10.1371/journal.pone.0005738
- Fang, F. C., Steen, R. G., & Casadevall, A. (2012). Misconduct accounts for the majority of retracted scientific publications. *Proceedings of the National Academy of Sciences of the United States of America*, 109, 17028–17033. doi:10.1073/pnas.1212247109
- Farley, H., Murphy, A., Johnson, C., Carter, B., Lane, M., Midgley, W., ... Koronios, A. (2015). How do students use their mobile devices to support learning? A case study from an Australian regional university. *Journal of Interactive Media in Education*, 2015(1), Article 14. doi:10.5334/jime.ar

- Field, A. P. (2010). Non-sadistical methods for teaching statistics. In D. Upton & A. Trapp (Eds.), *Teaching psychology in higher education* (pp. 134-163). Chester, England: Wiley-Blackwell. doi:10.1002/9781444320732.ch6
- Fok, L. Y., Angelidis, J. P., Ibrahim, N. A., & Fok, W. M. (1995). A decision tree approach to the interpretation of multivariate statistical techniques. *Journal of Education for Business, 71*, 110–117. doi:10.1080/08832323.1995.10116770
- Fox, R. J., Crask, M. R., & Kim, J. (1988). Mail survey response rate: A meta-analysis of selected techniques for inducing response. *Public Opinion Quarterly, 52*, 467-491. doi:10.1086/269125
- Franklyn-Stokes, A., & Newstead, S. E. (1995). Undergraduate cheating: Who does what and why? *Studies in Higher Education, 20*, 159–172. doi:10.1080/03075079512331381673
- Fredricks, J. A., Wang, M., Linn, J. S., Hofkens, T. L., Sung, H., Parr, A., & Allerton, J. (2016). Using qualitative methods to develop a survey measure of math and science engagement. *Learning and Instruction, 43*, 5-15. Doi:10.1016/j.learninstruc.2016.01.009
- Freeman, S., Eddy, S. L., McDonough, M., Smith, M. K., Okoroafor, N., Jordt, H., & Wenderoth, M. P. (2014). Active learning increases student performance in science, engineering, and mathematics. *Proceedings of the National Academy of Sciences of the United States of America, 111*, 8410–8415. doi:10.1073/pnas.1319030111
- Gaiser, T. J., & Schreiner, A. E. (2009). *A guide to conducting online research*. London, England: Sage.
- Gardner, P. L., & Hudson, I. (1999). University students' ability to apply statistical procedures. *Journal of Statistics Education, 7*(1). Retrieved from <http://www.amstat.org/publications/jse/secure/v7n1/gardner.cfm>
- Garfield, J., & Ahlgren, A. (1988). Difficulties in learning basic concepts in probability and statistics: Implications for research. *Journal of Research in Mathematics Education, 19*, 44–63. doi:10.2307/749110

- Garfield, J., & Ben-Zvi, D. (2007). How students learn statistics revisited: A current review of research on teaching and learning statistics. *International Statistical Review*, *75*, 372–396. doi:10.1111/j.1751-5823.2007
- Garfield, J., Hogg, B., Schau, C., & Whittinghill, D. (2002). First courses in statistical science: The status of educational reform efforts. *Journal of Statistics Education*, *10*(2). Retrieved from <http://www.amstat.org/publications/jse/v10n2/garfield.html>
- Gehlbach, H., & Brinkworth, M. E. (2011). Measure twice, cut down error: A process of enhancing the validity of survey scales. *Review of General Psychology*, *15*, 380-387. doi:10.1037/a0025704
- Gines, A. C. (2006). Teaching undergraduate psychology in the Philippines: A summary of current programs, policies, and instruction. *International Journal of Psychology*, *41*, 51-57. doi:10.1080/00207590444000465
- Goritz, A. S. (2010). Using lotteries, loyalty points, and other incentives to increase participant response and completion. In S. D. Gosling & J. A. Johnson (Eds.), *Advanced methods for conducting online behavioral research* (pp. 219-233). Washington, DC: American Psychological Association.
- Grahe, J. E., Reifman, A., Hermann, A. D., Walker, M., Oleson, K. C., Nario-Redmond, M., & Wiebe, R. P. (2012). Harnessing the undiscovered resource of student research projects. *Perspectives on Psychological Science*, *7*, 605-607. doi:10.1177/1745691612459057
- Grieneisen, M. L., & Zhang, M. (2012). A comprehensive survey of retracted articles from the scholarly literature. *PLoS One*, *7*, e44118. doi:10.1371/journal.pone.0044118
- Gross, C. (2016). Scientific misconduct. *Annual Review of Psychology*, *67*, 693-711. doi:10.1146/annurev-psych-122414-033437
- Guillemin, M., & Gillam, L. (2004). Ethics, reflexivity, and “ethically important moments” in research. *Qualitative Inquiry*, *10*, 261–280. doi:10.1177/1077800403262360

- Gurung, R. A. R. (2015). Conducting and applying the scholarship of teaching and learning. In D. S. Dunn (Ed.), *The Oxford handbook of undergraduate psychology education* (pp. 187-197). New York: Oxford University Press.
- Gurung, R. A. R., Ansbarg, P. L., Alexander, P. A., Lawrence, N. K., & Johnson, D. E. (2008). The state of scholarship of teaching and learning in psychology. *Teaching of Psychology, 35*, 249-261. doi:10.1080/00986280802374203
- Gurung, R. A. R., Hackathorn, J., Enns, C., Frantz, S., Cacioppo, J. T., Loop, T., & Freeman, J. E. (2016). Strengthening introductory psychology: A new model for teaching the introductory course. *American Psychologist, 71*, 112-124. doi:10.1037/a0040012
- Hamilton, M., & Geraci, L. (2004). Converting an experimental laboratory course from paper and pencil to computer. *Teaching of Psychology, 31*, 141-143. doi:10.1207/s15328023top3102_7
- Hanna, D., Shevlin, M., & Dempster, M. (2008). The structure of the statistics anxiety rating scale: A confirmatory factor analysis using UK psychology students. *Personality and Individual Differences, 45*, 68-74. doi:10.1016/j.paid.2008.02.021
- Harlow, L. L. (2013). Teaching quantitative psychology. In T. D. Little (Ed.), *The Oxford handbook of quantitative methods in psychology* (Vol 1, pp. 105-117). New York, NY: Oxford University Press. doi:10.1093/oxfordhb/9780199934874.013.0006
- Heerwegh, D., & Loosveldt, G. (2006). An experimental study on the effects of personalization, survey length statements, progress indicators, and survey sponsor logos in web surveys. *Journal of Official Statistics, 22*, 191-210. Retrieved from <http://www.jos.nu/Articles/abstract.asp?article=222191>
- Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences, 33*, 61-135. doi:10.1017/S0140525X0999152X
- Henry, P. J. (2008). College sophomores in the laboratory redux: Influence of a narrow data base on social psychology's view of the nature of prejudice. *Psychological Inquiry, 19*, 49-71. doi:10.1080/10478400802049936

- Higbee, K., Millard, R., & Folkman, J. (1982). Social psychology research during the 1970s: Predominance of experimentation and college students. *Personality and Social Psychology Bulletin*, *8*, 180-183. doi:10.1177/014616728281028
- Hogg, R. V. (1991). Statistical education: Improvements are badly needed. *American Statistician*, *45*, 342-343. doi:10.2307/2684473
- Holmes, J. D., & Beins, B. C. (2011). Student engagement in undergraduate research. In R. L. Miller, E. Amsel, B. M. Kowalewski, B. C. Beins, K. D. Keith, & B. F. Peden (Eds.), *Promoting student engagement* (Vol. 1, pp. 114-117). Retrieved from <http://teachpsych.org/ebooks/pse2011/index.php>
- Holmes, J. D., & Beins, B. C. (2015). Teaching laboratory courses in psychology. In D. S. Dunn (Ed.), *The Oxford handbook of undergraduate psychology education* (pp. 143-154). New York: Oxford University Press.
- James, N., & Busher, H. (2007). Ethical issues in online educational research: Protecting privacy, establishing authenticity in email interviewing. *International Journal of Research & Method in Education*, *30*, 101–113. doi:10.1080/17437270701207868
- John, L. K., Loewenstein, G., & Prelec, D. (2012). Measuring the prevalence of questionable research practices with incentives for truth telling. *Psychological Science*, *23*, 524–532. doi:10.1177/0956797611430953
- Kaczmirek, L. (2008). Internet survey software tools. In N. Fielding, R. M. Lee, & G. Blank (Eds.), *The SAGE handbook of online research methods* (pp. 236–254). London, England: Sage.
- Kardash, C. M. (2000). Evaluation of an undergraduate research experience: Perceptions of undergraduate interns and their faculty mentors. *Journal of Educational Psychology*, *92*, 191-201. doi:10.1037//0022-0663.92.1.191
- Karlins, M., Michaels, C., & Podlogar, S. (1988). An empirical investigation of actual cheating in a large sample of undergraduates. *Research in Higher Education*, *29*, 359–364. doi:10.1007/BF00992776
- Kays, K., Gathercoal, K., & Buhrow, W. (2012). Does survey format influence self-disclosure on sensitive question items? *Computers in Human Behavior*, *28*, 251-256. doi:10.1016/j.chb.2011.09.007

- Kidd, S. A. (2002). The role of qualitative research in psychological journals. *Psychological Methods, 7*, 126–138. doi:10.1037/1082-989X.7.1.126
- Kierniesky, N. C. (2005). Undergraduate research in small psychology departments: Two decades later. *Teaching of Psychology, 32*, 84–90. doi:10.1207/s15328023top3202_1
- Kim-Prieto, C., & D’Orlando, C. (2011) Integrating research training and the teaching of positive psychology. *The Journal of Positive Psychology, 6*, 457-462, doi:10.1080/17439760.2011.634827
- King, D. J. (1970). The subject pool. *American Psychologist, 25*, 1179-1181. doi:10.1037/h0037911
- Koch, C., & Gobell, J. (1999). A hypertext-based tutorial with links to the Web for teaching statistics and research methods. *Behavior Research Methods: Instruments & Computers, 31*, 7–13. doi:10.3758/BF03207686
- Korn, J. H. (1999). Students’ roles, rights, and responsibilities as research participants. In M. E. Ware & C. L. Brewer (Eds.), *Handbook for teaching statistics and research methods* (2nd ed.; pp. 113-118). Mahwah, NJ: Lawrence Erlbaum.
- Kretser, H., Wong, R., Robertson, S., Pershyn, C., Huang, J., Sun, F., ... Zahler, P. (2015). Mobile decision-tree tool technology as a means to detect wildlife crimes and build enforcement networks. *Biological Conservation, 189*, 33–38. doi:10.1016/j.biocon.2014.08.018
- Landrum, R. E., & Chastain, G. (1995). Experiment spot-checks: A method for assessing the educational value of undergraduate participation in research. *IRB: Ethics and Human Research, 17*, 4–6. doi:10.2307/3564152
- Landrum, R. E., & Nelsen, L. R. (2002). The undergraduate research assistantship: An analysis of the benefits. *Teaching of Psychology, 29*, 15-19. doi:10.1207/S15328023TOP2901_04
- Landrum, R. E., & Smith, R. A. (2007). Creating syllabi for statistics and research methods courses. In D. S. Dunn, R. A. Smith, & B. C. Beins (Eds.), *Best practices for teaching statistics and research methods in the behavioral sciences* (pp. 45-58). Mahwah, NJ: Lawrence Erlbaum Associates.

- Larkin, J. E., & Pines, H. A. (2005). Developing information literacy and research skills in introductory psychology: A case study. *Journal of Academic Librarianship*, *31*, 40-45. doi:10.1016/j.acalib.2004.09.008
- Lawson, A. E., Lewis, C. M. Jr., & Birk, J. P. (1999/2000). Why do students “cook” data? A case study on the tenacity of misconceptions. *Journal of College Science Teaching*, *29*, 191–198.
- Leak, G. K. (1981). Student perception of coercion and value from participation in psychological research. *Teaching of Psychology*, *8*, 147-149. doi:10.1207/s15328023top0803_4
- Lipp, O. V., Terry, D. J., Chalmers, D., Bath, D., Hannan, G., Martin, F., ... Provost, S. C. (2007). *Learning outcomes and curriculum development in psychology*. Sydney, Australia: Carrick Institute for Learning and Teaching in Higher Education. Retrieved from https://www.psy.uq.edu.au/carrick/downloads/Psychology_Final_Report.pdf
- Lipsitz, A. (2000). Research methods with a smile: A gender difference exercise that teaches methodology. *Teaching of Psychology*, *27*, 111–113. doi:10.1207/S15328023TOP2702_07
- Lovett, M. C., & Greenhouse, J. B. (2000). Applying cognitive theory to statistics instruction. *American Statistician*, *54*, 196-206. doi:10.2307/2685590
- Lozar Manfreda, K., & Vehovar, V. (2002, August). *Survey design features influencing response rates in web surveys*. Paper presented at the International Conference on Improving Surveys (ICIS-2002). Retrieved from http://www.websm.org/uploadi/editor/Lozar_Vehovar_2001_Survey_design.pdf
- Macher, D., Papousek, I., Ruggeri, K., & Paechter, M. (2015). Statistics anxiety and performance: Blessings in disguise. *Frontiers in Psychology*, *6*, Article 1116, doi:10.3389/fpsyg.2015.01116
- Madlock-Brown, C. R., & Eichmann, D. (2015). The (lack of) impact of retraction on citation networks. *Science and Engineering Ethics*, *21*, 127-137. doi:10.1007/s11948-014-9532-1

- Marchel, C., & Owens, S. (2007). Qualitative research in psychology: Could William James get a job? *History of Psychology, 10*, 301-324. doi:10.1037/1093-4510.10.4.301
- Marek, P., Christopher, A. N., & Walker, B. J. (2004). Learning by doing: Research methods with a theme. *Teaching of Psychology, 31*, 128–131. doi:10.1207/s15328023top3102_6
- Marsden, H., Carroll, M., & Neill, J. T. (2005). Who cheats at university? A self-report study of dishonest academic behaviours in a sample of Australian university students. *Australian Journal of Psychology, 57*, 1–10. doi:10.1080/00049530412331283426
- Martin, D. E., Rao, A., & Sloan, L. R. (2009). Plagiarism, integrity, and workplace deviance: A criterion study. *Ethics & Behavior, 19*, 36–50. doi:10.1080/10508420802623666
- Martin, F. H., Cranney, J., & Varcin, K. (2013). Students' experience of the psychology fourth year in Australia. *Proceedings of the Australian Conference on Science and Mathematics Education* (pp. 163-168). Retrieved from <http://openjournals.library.usyd.edu.au/index.php/IISME/article/view/7085>
- McCabe, D. L. (2005). Cheating among college and university students: A North American perspective. *International Journal of Educational Integrity, 1*(1). Retrieved from <http://www.ojs.unisa.edu.au/index.php/IJEI/article/view/14>
- McGhee, P. (2015, March 31). Let students be students – not customers. *The Guardian*. Retrieved from <https://www.theguardian.com/education/2015/mar/31/students-not-customers>
- McGovern, T. V., Corey, L., Cranney, J., Dixon, W. E. Jr., Holmes, J. D., Kuebli, J. E., ... Walker, S. J. (2010). Psychologically literate citizens. In D. F. Halpern (Ed.), *Undergraduate education in psychology: A blueprint for the future of the discipline* (pp. 9-27). Washington, DC: American Psychological Association. doi:10.1037/12063-001
- McNemar, Q. (1946). Opinion-attitude methodology. *Psychological Bulletin, 43*, 289-374. doi:10.1037/h0060985

- Messer, B. L., Edwards, M. L., & Dillman, D. A. (2012, April). Determinants of item nonresponse to web and mail respondents in three address-based mixed-mode surveys of the general public. *Survey Practice*. Retrieved from <http://surveypractice.wordpress.com/2012/04/17/determinants-of-item-nonresponse-to-web-and-mail-respondents-in-three-address-based-mixed-mode-surveys-of-the-general-public/>
- Messer, W. S., Griggs, R. A., & Jackson, S. L. (1999). A national survey of undergraduate psychology degree options and major requirements. *Teaching of Psychology, 26*, 164-171. doi:10.1207/S15328023TOP260301
- Miles, S. R., Cromer, L. D., & Narayan, A. (2015). Applying equity theory to students' perceptions of research participation requirements. *Teaching of Psychology, 42*, 349-356. doi:10.1177/0098628315603252
- Miller, R. L. (2015). Collaboration: Student-faculty research. In D. S. Dunn (Ed.), *The Oxford handbook of undergraduate psychology education* (pp. 225-240). New York: Oxford University Press.
- Miller, R. L., Rycek, R. F., Balcetis, E., Barney, S. T., Beins, B. C., Burns, S. R., ... Ware, M. E. (Eds.). (2008). *Developing, promoting, & sustaining the undergraduate research experience in psychology*. Washington, DC: Society for the Teaching of Psychology. Retrieved from <http://teachpsych.org/ebooks/ur2008/index.php>
- Moreland, R. L. (1999). Evaluating students' research experiences via credit slips. In G. Chastain & R. E. Landrum (Eds.), *Protecting human subjects: Departmental subject pools and institutional review boards* (pp. 87-108). Washington, DC: American Psychological Association.
- Morgan, D. L. (2009). Using single-case design and personalized behavior change projects to teach research methods. *Teaching of Psychology, 36*, 267-269. doi:10.1080/00986280903175715
- Moyer, A., & Franklin, N. (2011). Strengthening the educational value of undergraduate participation in research as part of a psychology department subject pool. *Journal of Empirical Research on Human Research Ethics, 6*, 75-82. doi:10.1525/jer.2011.6.1.75

- Munley, P. H., Anderson, M. Z., Briggs, D., Devries, M. R., Forshee, W. J., & Whisner, E. A. (2002). Methodological diversity of research published in selected psychological journals in 1999. *Psychological Reports, 91*, 411–420. doi:10.2466/pr0.2002.91.2.411
- Murray, D. M., & Hannan, P. J. (1990). Planning for the appropriate analysis in school-based drug-use prevention studies. *Journal of Consulting and Clinical Psychology, 58*, 458-468. doi:10.1037/0022-006X.58.4.458
- Murtonen, M. (2005). University students' research orientations: Do negative attitudes exist toward quantitative methods? *Scandinavian Journal of Educational Research, 49*, 263-280. doi:10.1080/00313830500109568
- Murtonen, M. (2015). University students' understanding of the concepts empirical, theoretical, qualitative and quantitative research. *Teaching in Higher Education, 20*, 684-698. doi:10.1080/13562517.2015.1072152
- Murtonen, M., & Lehtinen, E. (2003). Difficulties experienced by education and sociology students in quantitative methods courses. *Studies in Higher Education, 28*, 171–185. doi:10.1080/0307507032000058064
- Murtonen, M., Olkinuora, E., Tynjala, P., & Lehtinen, E. (2008). “Do I need research skills in working life?”: University students' motivation and difficulties in quantitative methods courses. *Higher Education, 56*, 599–612. doi: 10.1007/s10734- 008-9113-9
- Neumann, D. L., Hood, M., & Neumann, M. M. (2013). Using real-life data when teaching statistics: Student perceptions of this strategy in an introductory statistics course. *Statistics Education Research Journal, 12*, 59–70. Retrieved from [http://iase-web.org/documents/SERJ/SERJ12\(2\)_Neumann.pdf](http://iase-web.org/documents/SERJ/SERJ12(2)_Neumann.pdf)
- Neumann, D. L., Neumann, M. M., & Hood, M. (2010). The development and evaluation of a survey that makes use of student data to teach statistics. *Journal of Statistics Education, 18*, 1–19. Retrieved from <http://www.amstat.org/publications/jse/v18n1/neumann.pdf>
- Nichter, M., Nichter, M., Thompson, P. J., Shiffman, S., & Moscicki, A. (2002). Using qualitative research to inform survey development on nicotine dependence

- among adolescents. *Drug and Alcohol Dependence*, 68, S41-S56. doi:10.1016/S0376-8716(02)00214-4
- Nimmer, J. G., & Handelsman, M. M. (1992). Effects of subject pool policy on student attitudes toward psychology and psychological research. *Teaching of Psychology*, 19, 141-144. doi:10.1207/s15328023top1903_2
- Nonis, S., & Swift, C. O. (2001). An examination of the relationship between academic dishonesty and workplace dishonesty: A multicampus investigation. *Journal of Education for Business*, 77, 69-77. doi:10.1080/08832320109599052
- Norcross, J. C., Hailstorks, R., Aiken, L. S., Pfund, R. A., Stamm, K. E., & Christidis, P. (2016). Undergraduate study in psychology: Curriculum and assessment. *American Psychologist*, 71, 89-101. doi:10.1037/a0040095
- Nosek, B. A., Sriram, N., & Umansky, E. (2012). Presenting survey items one at a time compared to all at once decreases missing data without sacrificing validity in research with Internet volunteers. *PLoS One*, 7, e36771. doi:10.1371/journal.pone.0036771
- O'Neil, K. M., Penrod, S. D., & Bornstein, B. H. (2003). Web-based research: Methodological variables' effects on dropout and sample characteristics. *Behavior Research Methods, Instruments, & Computers*, 35, 217-226. doi:10.3758/BF03202544
- Onwuegbuzie, A. J., & Wilson, V. A. (2003). Statistics anxiety: Nature, etiology, antecedents, effects, and treatments: A comprehensive review of the literature. *Teaching in Higher Education*, 8, 195-209. doi:10.1080/1356251032000052447
- Perlman, B., & McCann, L. I. (1999a). The most frequently listed courses in the undergraduate psychology curriculum. *Teaching of Psychology*, 26, 177-182. doi:10.1207/S15328023TOP260303
- Perlman, B., & McCann, L. I. (1999b). The structure of the psychology undergraduate curriculum. *Teaching of Psychology*, 26, 171-176. doi:10.1207/S15328023TOP260302

- Perlman, B., & McCann, L. I. (2005). Undergraduate research experiences in psychology: A national study of courses and curricula. *Teaching of Psychology, 32*, 5–14. doi:10.1207/s15328023top3201_2
- Peterson, R. A. (1975). An experimental investigation of mail-survey responses. *Journal of Business Research, 3*, 199-210. doi:10.1016/0148-2963(75)90022-3
- Protsman, L., & Carlson, M. (2008). Graphic organizers can facilitate selection of statistical tests: Part 2 – Correlation and regression analysis. *Journal of Physical Therapy Education, 22*, 36–41. Retrieved from <http://aptaeducation.org/members/jopte/index.cfm>
- Public Health Service Policies on Research Misconduct, 93 C.F.R. §93.103 (2005).
- Pullen, R., Ortloff, V., Casey, S., & Payne, J. B. (2000). Analysis of academic misconduct using unobtrusive research: A study of discarded cheat sheets. *College Student Journal, 34*, 616–625.
- Quality Assurance Agency for Higher Education. (2010). *Subject benchmark statement – Psychology: 2007* (3rd ed.). Gloucester, England: Author. Retrieved from <http://www.qaa.ac.uk/en/Publications/Documents/Subject-benchmark-statement-Psychology.pdf>
- Quilici, J. L., & Mayer, R. E. (2002). Teaching students to recognize structural similarities between statistics word problems. *Applied Cognitive Psychology, 16*, 325–342. doi:10.1002/acp.796
- Rabinowitz, M., & Hogan, T. M. (2008). Experience and problem representation in statistics. *American Journal of Psychology, 121*, 395–407. doi:10.2307/20445474
- Ragozzine, F. (2002). SuperLab LT: Evaluation and uses in teaching experimental psychology. *Teaching of Psychology, 29*, 251–254. doi:10.1207/S15328023TOP2 903_13
- Rajah-Kanagasabai, C. J., & Roberts, L. D. (2015). Predicting self-reported research misconduct and questionable research practices in university students using an augmented Theory of Planned Behavior. *Frontiers in Psychology, 6*, Article 535. doi:10.3389/fpsyg.2015.00535

- Ranstam, J., Buyse, M., George, S. L., Evans, S., Geller, N. L., Scherrer, B. ... Lachenbruch, P. (2000). Fraud in medical research: An international survey of biostatisticians. *Controlled Clinical Trials*, *21*, 415-427. doi:10.1016/S0197-2456(00)00069-6
- Rennie, D. L., Watson, K. D., & Monteiro, A. M. (2002). The rise of qualitative research in psychology. *Canadian Psychology/Psychologie Canadienne*, *43*, 179–189. doi:10.1037/h0086914
- Roberts, L. D. (2015a). *Best practice: Honours and coursework dissertation supervision*. Sydney, Australia: Office for Learning and Teaching. doi:10.13140/RG.2.1.4974.4720
- Roberts, L. D. (2015b). *Guide for new supervisors of honours and coursework dissertation students*. Retrieved from <http://www.dissertationsupervision.org>
- Roberts, L. D., & Allen, P. J. (2012). Student perspectives on the value of research participation. In S. McCarthy, K. L. Dickson, J. Cranney, A. Trapp, & V. Karandashev (Eds.), *Teaching psychology around the world* (Vol 3; pp. 198-211). Newcastle, England: Cambridge Scholars.
- Roberts, L. D., & Allen, P. J. (2013). A brief measure of student perceptions of the educational value of research participation. *Australian Journal of Psychology*, *65*, 22-29. doi:10.1111/ajpy.12007
- Roberts, L. D., & Allen, P. J. (2015). Exploring ethical issues associated with using online surveys in educational research. *Educational Research and Evaluation*, *21*, 95-108. doi:10.1080/13803611.2015.1024421
- Roberts, L. D., Heritage, B., & Gasson, N. (2015). The measurement of psychological literacy: A first approximation. *Frontiers in Psychology*, *6*, Article 105. doi:10.3389/fpsyg.2015.00105
- Rottinghaus, P. J., Gaffey, A. R., Borgen, F. H., & Ralston, C. A. (2006). Diverse pathways of psychology majors: Vocational interests, self-efficacy, and intentions. *The Career Development Quarterly*, *55*, 85-93. doi:10.1002/j.2161-0045.2006.tb00007.x

- Rowan, N., & Wulff, D. (2007). Using qualitative methods to inform scale development. *The Qualitative Report*, *12*, 450-466. Retrieved from <http://www.nova.edu/ssss/QR/QR12-3/rowan.pdf>
- Sanchez-Fernandez, J., Munoz-Leiva, F., & Montoro-Rios, F. J. (2012). Improving retention rate and response quality in web-based surveys. *Computers in Human Behavior*, *28*, 507-514. doi:10.1016/j.chb.2011.10.023
- Saville, B. K. (2015). Teaching research methods. In D. S. Dunn (Ed.), *The Oxford handbook of undergraduate psychology education* (pp. 323-341). New York: Oxford University Press.
- Schau, C., & Mattern, N. (1997). Use of map techniques in teaching applied statistics courses. *American Statistician*, *51*, 171-175. doi:10.1080/00031305.1997.10473955
- Schoenherr, J. R. (2015). Scientific integrity in research methods. *Frontiers in Psychology*, *6*, Article 1562. doi:10.3389/fpsyg.2015.01562
- Sears, D. O. (1986). College sophomores in the laboratory: Influences of a narrow data base on social psychology's view of human nature. *Journal of Personality and Social Psychology*, *51*, 515-530. doi:10.1037/0022-3514.51.3.515
- Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston, MA: Houghton Mifflin.
- Sharrock, G. (2013, May 9) Students aren't customers ... or are they? *The Conversation*. Retrieved from <http://theconversation.com/students-arent-customers-or-are-they-13282>
- Sieber, J. E. (1999). What makes a subject pool (un)ethical? In G. D. Chastain & R. E. Landrum (Eds.), *Protecting human subjects: Departmental subject pools and institutional review boards* (pp. 43- 68). Washington, DC: American Psychological Association.
- Sieber, J. E., & Saks, M. J. (1989). A census of subject pool characteristics and policies. *American Psychologist*, *44*, 1053-1061. doi:10.1037/0003-066X.44.7.1053

- Simons, H., & Usher, R. (Eds.). (2000). *Situated ethics in education research*. New York, NY: Routledge/Falmer.
- Singer, J. D., & Willett, J. B. (1990). Improving the teaching of applied statistics: Putting the data back into data analysis. *American Statistician*, *44*, 223–230. doi:10.2307/2685342
- Sizemore, O. J., & Lewandowski, G. W., Jr., (2009). Learning might not equal liking: Research methods course changes knowledge but not attitudes. *Teaching of Psychology*, *36*, 90-95. doi:10.1080/00986280902739727
- Stedman, M. E. (1993). Statistical pedagogy: Employing student-generated data sets in introductory statistics. *Psychological Reports*, *72*, 1036–1038. doi:10.2466/pr0.1993.72.3.1036
- Stieger, S., Reips, U. –D., & Voracek, M. (2007). Forced-response in online surveys: Bias from reactance and an increase in sex-specific dropout. *Journal of the American Society for Information Science and Technology*, *58*, 1653-1660. doi:10.1002/asi.20651
- Steen, R. G. (2011). Retractions in the scientific literature: Is the incidence of research fraud increasing? *Journal of Medical Ethics*, *37*, 249–253. doi:10.1136/jme.2010.040923
- Steneck, N. H. (2006). Fostering integrity in research: Definitions, current knowledge and future directions. *Science and Engineering Ethics*, *12*, 53–74. doi:10.1007/PL00022268
- Stoloff, M., McCarthy, M., Keller, L., Varfolomeeva, V., Lynch, J., Makara, K., ... Smiley, W. (2010). The undergraduate psychology major: An examination of structure and sequence. *Teaching of Psychology*, *37*, 4-15. doi:10.1080/00986280903426274
- Stoloff, M. L., Curtis, N. A., Rodgers, M., Brewster, J., & McCarthy, M. A. (2012). Characteristics of successful undergraduate psychology programs. *Teaching of Psychology*, *39*, 91-99. doi:10.1177/0098628312437721
- Stoloff, M. L., Good, M. R., Smith, K. L., & Brewster, J. (2015). Characteristics of programs that maximize psychology major success. *Teaching of Psychology*, *42*, 99–108. doi:10.1177/0098628315569877

- Stone, T. H., Jawahar, I. M., & Kisamore, J. L. (2011). Predicting workplace misconduct using personality and academic behaviors. In R. J. Burke, E. C. Tomlinson, & C. L. Cooper (Eds.), *Crime and corruption in organizations: Why it occurs and what to do about it* (pp. 97-119). London, England: Taylor & Francis.
- Stroebe, W., Postmes, T., & Spears, R. (2012). Scientific misconduct and the myth of self-correction in science. *Perspectives on Psychological Science*, 7, 670–688. doi:10.1177/1745691612460687
- Sue, V. M., & Ritter, L. A. (2007). *Conducting online surveys*. Thousand Oaks, CA: Sage.
- Sumer, N. (2016). Rapid growth of psychology programs in Turkey: Undergraduate curriculum and structural challenges. *Teaching of Psychology*, 43, 63-69. doi:10.1177/0098628315620886
- Swazey, J. P., Anderson, M. S., & Louis, K. S. (1993). Ethical problems in academic research. *American Scientist*, 81, 542–553.
- Tabachnick, B. G., & Fidell, L. S. (2013). *Using multivariate statistics* (6th ed.). Boston, MA: Pearson.
- Tabachnick, B. G., Keith-Spiegel, P., & Pope, K. S. (1991). Ethics of teaching: Beliefs and behaviors of psychologists as educators. *American Psychologist*, 46, 506-515. doi:10.1037/0003-066X.46.5.506
- Thompson, W. B. (1994). Making data analysis realistic: Incorporating research into statistics courses. *Teaching of Psychology*, 21, 41–43. doi:10.1207/s15328023top2101_9
- Titus, S. L., Wells, J. A., & Rhoades, L. J. (2008). Repairing research integrity. *Nature*, 453, 980-982. doi:10.1038/453980a
- Trafimow, D., Madson, L., & Gwizdowski, L. (2006). Introductory psychology students' perceptions of alternatives to research participation. *Teaching of Psychology*, 33, 247–249. doi:10.1207/s15328023top3304_7
- Twycross, A., & Shields, L. (2004). Statistics made simple: Part 4 – Choosing the right statistical test. *Paediatric Nursing*, 16, 24. doi:10.7748/paed2004.10.16.8.24.c944

- VanWormer, L. A., Jordan, E. F., & Blalock, L. D. (2014). Assessing the perceived value of research participation. *Teaching of Psychology, 41*, 233-236. doi:10.1177/0098628314537974
- Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. *MIS Quarterly, 27*, 425–478.
- Vespia, K. M., Wilson-Doenges, G., Martin, R. C., & Radosevich, D. M. (2012). Experiential learning. In B. M. Schwartz & R. A. R. Gurung (Eds.), *Evidence-based teaching for higher education* (pp. 77-97). Washington, DC. American Psychological Association.
- Vittengl, J. R., Bosley, C. Y., Brescia, S. A., Eckardt, E. A., Neidig, J. M., Shelver, K. S., & Sopenoff, L. A. (2004). Why are some undergraduates more (and others less) interested in psychological research? *Teaching of Psychology, 31*, 91-97. doi:10.1207/s15328023top3102_3
- Wagner, E., & Williams, P. (2011). Why and how do journals retract articles? An analysis of Medline retractions 1988–2008. *Journal of Medical Ethics, 37*, 567-570. doi:10.1136/jme.2010.040964
- Ward, D. A., & Beck, W. L. (1990). Gender and dishonesty. *The Journal of Social Psychology, 130*, 333–339. doi:10.1080/00224545.1990.9924589
- Ware, M. E., & Chastain, J. D. (1989). Computer-assisted statistical analysis: A teaching innovation? *Teaching of Psychology, 16*, 222–227. doi:10.1207/s15328023top1604_16
- Ware, M. E., & Chastain, J. D. (1991). Developing selection skills in introductory statistics. *Teaching of Psychology, 18*, 219–222. doi:10.1207/s15328023top1804_4
- Wayment, H. A., & Dickson, K. L. (2008). Increasing student participation in undergraduate research benefits students, faculty, and department. *Teaching of Psychology, 35*, 194-197. doi:10.1080/00986280802189213
- Wells, J. A. (2008). *Final report: Observing and reporting suspected misconduct in biomedical research*. Washington, DC: Gallup Organization. Retrieved from http://ori.hhs.gov/sites/default/files/gallup_finalreport.pdf

- Williams, J., & Roberts, D. (2016). *Academic integrity: Exploring the tensions between perception and practice in the contemporary university*. London, England: Society for Research into Higher Education. Retrieved from <http://www.srhe.ac.uk/downloads/WILLIAMSJoannaROBERTSDavid.pdf>
- Wilson, P., & Provost, S. (2006). Psychology in Australian universities. *International Journal of Psychology, 41*, 3-9. doi:10.1080/00207590444000393
- Wilson, S. G. (2013). The flipped class: A method to address the challenges of an undergraduate statistics course, *Teaching of Psychology, 40*, 193-199. doi:10.1177/0098628313487461
- Wilson-Doenges, G., & Gurung, R. A. R. (2013). Benchmarks for scholarly investigations of teaching and learning. *Australian Journal of Psychology, 65*, 63-70. doi:10.1111/ajpy.12011
- Wilson-Doenges, G., Troisi, J. D., & Bartsch, R. A. (2016). Exemplars of the gold standard in SoTL for psychology. *Scholarship of Teaching and Learning in Psychology, 2*, 1-12. doi:10.1037/stl0000050
- Wintre, M. G., North, C., & Sugar, L. A. (2001). Psychologists' response to criticisms about research based on undergraduate participants: A developmental perspective. *Canadian Psychology/Psychologie Canadienne, 42*, 216-225. doi:10.1037/h0086893
- Woodzicka, J. A., Ford, T. E., Caudill, A., & Ohanmamooreni, A. (2015). A successful model of collaborative undergraduate research: A multi-faculty, multi-project, multi-institution team approach. *Teaching of Psychology, 42*, 60-63. doi:10.1177/0098628314549711
- Wright, K. B. (2005). Researching Internet-based populations: Advantages and disadvantages of on- line survey research, online questionnaire authoring software packages, and web survey services. *Journal of Computer-Mediated Communication, 10*(3). Retrieved from <http://jcmc.indiana.edu/vol10/issue3/wright.html>
- Yan, J., & Lavigne, N. C. (2014). Promoting college students' problem understanding using schema-emphasizing worked examples. *Journal of Experimental Education, 82*, 74-102. doi:10.1080/00220973.2012.745466

- Yan, T., Conrad, F. G., Tourangeau, R., & Couper, M. P. (2011). Should I stay or should I go: The effects of progress feedback, promised task duration, and length on questionnaire on completing web surveys. *International Journal of Public Opinion Research*, 23, 131-147. doi:10.1093/ijpor/edq046
- Yang, S. C. (2012) Attitudes and behaviors related to academic dishonesty: A survey of Taiwanese graduate students, *Ethics & Behavior*, 22, 218-237. doi:10.1080/10508422.2012.672904
- Yin, Y. (2012). Using tree diagrams as an assessment tool in statistics education. *Educational Assessment*, 17, 22–49. doi:10.1080/10627197.2012.697850

Every reasonable effort has been made to acknowledge the owners of copyright material. I would be pleased to hear from any copyright owner who has been omitted or incorrectly acknowledged.

Paper 1: Student Perspectives on the Value of Research Participation

Roberts, L. D., & Allen, P. J. (2012). Student perspectives on the value of research participation. In S. McCarthy, K. L. Dickson, J. Cranney, A. Trapp, & V. Karandashev (Eds.), *Teaching psychology around the world* (vol 3, pp. 198-211). Newcastle, England: Cambridge Scholars.

This is the post-print version of the abovementioned work. Readers wishing to cite this paper are encouraged to source the final published version, available from Cambridge Scholars.

Student Perspectives on the Value of Research Participation

Lynne D. Roberts and Peter J. Allen

Curtin University

Author Note

Lynne D. Roberts, School of Psychology and Speech Pathology, Curtin Health Innovation Research Institute, Curtin University; Peter J Allen, School of Psychology and Speech Pathology, Curtin Health Innovation Research Institute, Curtin University.

We would like to acknowledge the work of Erin Smith, Nyree Gregson and Suevani Naidoo in transcribing the exam answers for this research.

Correspondence concerning this chapter should be addressed to Dr Lynne D. Roberts, Senior Lecturer, School of Psychology and Speech Pathology, Curtin Health Innovation Research Institute, Curtin University, GPO Box U1987, Perth, Western Australia, 6845. E-mail: Lynne.Roberts@curtin.edu.au

Abstract

Undergraduate (UG) students are the most frequently used participants in psychological research. Here we report on the results of a qualitative exploration of the research participation experience, as seen from the perspective of UG psychology students. Following retrospective 'opt out' consent procedures, 143 first and third year psychology students' responses to a research methods exam question, "You've been invited to participate in a number of research projects this semester; what have you learned from this experience?" were de-identified, transcribed, and thematically analysed. The results provide a rich, nuanced and contextually sensitive account of the perceived value of the research participation experience. The major theme to emerge was that participating in research provides psychology students with increased insight into the research process. We propose that this educational gain may be further enhanced through greater integration of research participation with the teaching of research methods.

Student Perspectives on the Value of Research Participation

In 1946, McNemar quipped that “the existing science of human behavior is largely the science of the behavior of sophomores” (p. 333). Although much has changed in the intervening half-century, undergraduate (UG) students have remained as one of psychology’s favourite sampling frames. For example, Wintre, North, and Sugar (2001) reported that 68% of articles published in six major psychology journals representing five subdivisions of psychology during 1995 were based on UG student research participants. This figure has remained virtually unchanged since 1975 (Wintre et al., 2001). Within sub-disciplines of psychology, the prevalence of the use of student samples varies, with over 70% in personality and social psychology (Higbee, Millard, & Folkman, 1982; Korn, 1999; Sieber & Saks, 1989) and closer to 90% in the areas of perception and cognition (Korn, 1999).

The recruitment of UG students for psychological research participation commonly occurs through the use of “subject” or participant pools, where UG psychology students are required to participate in research (or other activities deemed equivalent) as a course or unit requirement. Where research participation is not mandated, incentives such as extra credit may be offered (Sieber & Saks, 1989). Internationally, estimates of the percentage of psychology schools/departments that have a participant pool range from 44% in the United Kingdom (Coulson, 1999) to 67% in Canada (Lindsay & Holden, 1987), 68% in Australia (Diamond & Reidpath, 1992) and 74% in the United States (Sieber & Saks, 1989; in US universities without postgraduate programs, the figure is somewhat lower at 32.7%; Landrum & Chastain, 1999).

This continued reliance on psychology UG students as research participants has been criticised on a number of grounds. First, it is questionable whether the results of research based on psychology UG students are generalizable to a wider population, as UG students differ from the wider population in ways that may introduce systematic biases into research findings (Henrich, Heine, & Norenzayan, 2010; Norenzayan & Heine, 2005; Sears, 1986). A second concern with the use of UG students is the dual relationships that exist when academics conduct research using their own students (Clark & McCann, 2005; Ferguson, Myrick, & Yonge, 2006; Ferguson, Yonge, & Myrick, 2004; Shi, 2006). A third area of concern is the potential for, or perception of, coercion when students are required to participate in research as a course requirement or for extra credit (Midzinski, 2010; Miller & Kreiner, 2008). Research participation for course credit may affect both responding and perceptions of freedom to withdraw from studies (Fisher & Fryberg, 1994). While some psychologists have ethical concerns about research participation being specified as a course requirement (Tabachnick, Keith-Spiegel, & Pope, 1991), the perception of required participation as coercive is certainly not shared by all. For instance, Dalziel (1996) noted that required research participation is no more coercive than other accepted student requirements, including exams and essays. Instead, Dalziel argued that the question is not one of whether research participation is coercive, but whether it is educationally justifiable.

Ethical research balances potential benefits from research against potential risks to research participants and others (National Health and Medical Research Council [NH&MRC], Australian Research Council [ARC], & Australian Vice-Chancellors' Committee [AVCC], 2007). From this perspective, there are clear benefits to society

(valuable research can be conducted) and to researchers (easy access to research participants) from student research participation. But what are the potential risks and benefits of research participation for the students participating? Is research participation educationally and ethically justifiable?

Existing research suggests that while the majority of students do not experience any form of harm, discomfort or inconvenience from research participation, a minority do experience some distress (Britton, Richardson, Smith, & Hamilton, 1983; Daugherty & Lawrence, 1996; Fligel, Best, & Hunter, 2007). However, this distress is usually mild (Fligel et al., 2007), and rarely enduring (Britton et al., 1983).

When examining possible the benefits of research participation for students, the focus has been on educational value. Student participation in research has traditionally been conceptualised as a form of “experiential learning” or “professional socialisation” (Clark & McCann, 2005, p. 42), with recognition that traditional lectures and texts are limited when conveying research procedures (Kimble, 1987). Specific educational values theorised include exposure to a range of psychological concepts and research methods, development of critical thinking when comparing research projects and building a basis for developing a student’s own research capabilities (Dalziel, 1996; Moreland, 1999). However, to obtain educational benefits from research participation, the ways in which the research is related to course outcomes need to be explicit. Unfortunately, all too many of the research projects students are required to participate in are either not directly relevant to course learning outcomes, or are not perceived as such by student participants (Coulter, 1986; Davis & Fernald, 1975; Diamond & Reidpath, 1992).

The majority of psychology schools and departments make no effort to evaluate the educational value of research participation for students (Landrum & Chastain, 1999). Published research to date that has examined the educational value of research participation has largely been quantitative and survey based. Survey findings indicate that some, but not all students see research participation as having some educational value, with endorsement of items relating to overall education value, learning about psychology, understanding research and increasing interest in psychology (Davis & Fernald, 1975; Landrum & Chastain, 1995; Leak, 1981; Moreland 1999; Trafimow, Madson, & Gwizdowski, 2006). Key factors identified as impacting on the educational value of research participation are the degree of explanation/debriefing provided (King, 1970) and the amount of research participated in, with judgements of the educational value of research participation decreasing as the number of studies participated in increased over time (Nimmer & Handelsman, 1992).

Only one true experiment in this area has examined educational gains. Elliott, Rice, Trafimow, Madson, and Hipshur (2010) randomly assigned students to a lecture on chunking in memory or participation in a chunking experiment followed by debriefing. While there was a main effect for knowledge gained from pre-test to post-test, there was no significant difference across conditions, leading Elliott and colleagues to conclude that learning across conditions was “approximately equal” (p. 130). The results of this study suggest that under ideal conditions students can gain educational benefits from research participation. However, the ecological validity of this study is questionable when one considers the range of conditions (e.g., extent of debriefing provided, individual differences

between researchers) that might exist within the range of research projects that draw from a typical participant pool.

There is limited published qualitative research that has asked student research participants about what they have learned as a result of research participation. Brody, Gluck, and Aragon (2000) interviewed 65 undergraduates from an UG participant pool. Less than a third (32%) rated the experience as positive and none mentioned any educational value from participation. Less than half (40.6%) considered that their debriefing was performed well. While the majority (63%) were positive about psychological research and the benefits to society, a substantial minority were not. Brody and colleagues reported that students felt “disengaged” by their participation, questioning the relevance of research and highlighting benefits to researchers at the expense of research participants.

A number of suggestions for how the educational value of research participation can be improved have been made. These can be divided into actions by researchers and actions by educators. Actions by researchers include mandatory debriefing immediately after participation (Coulter, 1986; Davis & Fernald, 1975; Sieber, 1999) as well as later feedback (Dalziel, 1996) through posters displays (Diamond & Reidpath, 2002) and written feedback reports (Coren, 1987). Actions by educators include requiring students to prepare reports on the projects they’ve participated in (Davis & Fernald, 1975; Richardson, Pegalis, & Britton, 1992) and integrating teaching and research participation through the linking of discussions of theories, methods (Dalziel, 1996) and research ethics (Sullivan & Lashley, 2009) to available research projects.

In summary, the existing research suggests there is educational value in research participation for some students under some conditions, and that there are actions that can be undertaken by both researchers and educators to potentially increase its educational value. However, the limited published research on the educational benefits of student research participation beyond the measurement of a few items on surveys suggests that more research is required. In particular, there is an absence of knowledge about what exactly students think they learn from participation in research. As noted by Landrum and Chastain (1999):

More scholarly research needs to be conducted in these areas. If we consider the number of participant-hours students provide to researchers nationwide, it seems clear that we have an obligation to ensure that the experience is educational, non-stressful and occurs without coercion. To do anything less would be unethical. (p. 40)

In this chapter we present the results of a qualitative study examining students' views of voluntary participation in research.

Method

Design

This research utilised a qualitative research design involving thematic analysis of existing textual data.

Participants

The starting sample for this research was 170 first year and 48 third year UG psychology students who sat an end of semester research methods examination in second semester 2008 at Curtin University in Perth, Western Australia, and were required (first

year students) or elected (third year students) to answer the question: “You’ve been invited to participate in a number of research projects this semester. What have you learned from this experience?” Twenty-three students from the starting sample were excluded as they opted out of the research, were not contactable by the researchers, or were first year students who did not answer the exam question. A further 55 cases were identified as being of no further interest in the analysis as they indicated they did not participate in any research or provided answers that did not appear to relate to research participation, typically reciting information learned in research methods classes. The remaining 143 transcripts were retained for analysis.

Procedure

We teach research methods to UG psychology students, and strongly encourage them to participate in research projects undertaken by fourth year and post-graduate students at our university. However, research participation is not a course requirement and no extra credit is provided for participating in research. In late 2008 first and third year students completing research methods units we teach had the question above included in their examination papers. As a way of encouraging students to participate in research, students were advised at the start of the semester that this question would be included in the examination.

On marking this exam question, we noted that students were providing some interesting and thoughtful answers. In the beginning of 2009, retrospective ethics approval from our Human Research Ethics Committee was approved to analyse the 2008 answers as a research project with opt out consent procedures. Answers to the question were deidentified and transcribed, with 15% of transcripts checked against the original exam

papers for accuracy of transcription. The transcripts were imported into NVivo 8.0 for analysis using the thematic analysis procedures outlined by Braun and Clarke (2006).

Results

The major theme to emerge from the analysis was that research participation provided insight into the research process. The experience of participation enhanced students' knowledge about the conduct of psychological research. This theme, insight into the research process, was expressed in seven subthemes, each of which is outlined below, and illustrated with quotes from students' exam answers.

Varieties of Psychological Research

Students noted that participating in research had provided insight into the range of topics and issues that were studied by psychologists, for example: "I found this experience informed my concept of the type of study psychology could be applied to, the applications psychology studies could be implemented in...". Further, students noted that participating in a range of studies had raised their awareness of the type of research conducted within their own university, "It also made me aware of the current research happening at Curtin", and "Participating in the research projects was more interesting than I thought. For one, the topics and areas being researched were actually interesting. I learnt that not all research had to be boring and that students seem to have been granted a certain amount of creative licence when choosing a project."

How 'Real' Research is Conducted

Second, insight into the research process was sometimes expressed in terms of how "real" research (as opposed to the theoretical material and practice exercises provided in research methods units) is conducted. This was described variously as "How to put research

into practice”, “Participating in research projects has given me insight as to how real researchers conduct experiments”, and “Being aware of the process which goes into research data has added a dimension to understanding what this data means”.

Role of the Researcher

Some participants described what they had learned about the role of the researcher, “...how the research experimenters behave and act throughout the projects”. The social skills of the researcher, “Human interaction is far more important than I thought it was”, and the impact they could have on participants’ comfort levels during research were highlighted, “Build a rapport - make participant comfortable”, and “Be aware of the body language and responses, not to cause experimenter bias, or performance anxiety on participant”.

Gaining Participant Perspective of Research

Students also commented on the opportunity participation provided to view research from a participant’s perspective: “It also gave me a chance to see research from the point of view of a participant, rather than a researcher.” Students reflected on how this knowledge could be used to inform their own research: “As I think about doing research someday, I find myself evaluating my experience as a participant. Does this sort of question work? Do I want to continue with this study? These questions have a practical significance informing my perspective on how to do successful research in the future.”

Process Issues

Some exam question responses related to insights about process issues in conducting research. These included comments on the complexity of conducting research, “A lot more goes on behind the scenes than we realise, research, if done thoroughly is a

long tedious experience unlike the small tutorial we did last semester in [first year research methods unit]”, and the commitment required, “Makes one aware of the copious time and energy involved in generating and following through with an experiment”.

Questionnaire Design

Some students reported specific, rather than global learning that would inform their future research endeavours. Most commonly these related to questionnaire design, covering everything from questionnaire length and presentation to item wording: “I came across questions which I thought were intrusive, worded wrongly, boring questions. I was able to learn between which ways I think were appropriate and how in future I may word questions or logically set out a questionnaire.” Responses suggested that students were reflecting on their participation experience to think about how they could improve on the research they had participated in: “When addressing these problems myself taking precautions to avoid these leading questions will represent truer results and increase the chances of successfully achieving solutions or understandings.”

Future Research Ideas

Some students noted that the insights into research obtained could be used in their own future research. For first year students, this would still be a few years away: “It has given me first-hand experience in research, and showed me what I will be doing in later years of my course”. For third year students, the prospect of conducting their own research the following year provided a more immediate perspective: “...allowed me to better prepare myself for the uphill task of conducting my own research project next year”.

While factors affecting research participation was a major theme to emerge from the research, a small number of students offered alternative views. Not all students reported

insight into the research process through research participation. Two first year students commented: “To be honest I have not learned that much”, and “I have ... to be honest, not learned a lot in relation to participating in the research projects”. In addition, some students described the experience as boring, tedious, anxiety provoking or intimidating. However, the majority reported the experience to be positive, using terms like “enjoyable”, “interesting”, “engaging” and “educational”.

Discussion

The aim of this research was to examine psychology students’ views of voluntary participation in research. The major theme to emerge from this research was that participating in research increased students’ insight into the research process. This included enhancing their knowledge of the types of psychological research conducted, providing insight into how ‘real’ research is conducted, the role of the researcher, process and questionnaire design issues. In addition, participating in research provided students with opportunities to gain a participant’s perspective of research against which they could compare the “researcher’s perspective” offered in class exercises. Our participants saw these insights as being directly applicable to their own future research.

In interpreting these findings, it is useful to keep in mind that both the major strength and the major limitation of this study was the use of exam answers as data. Because we were working with exam answers to a compulsory question in a compulsory unit in the first year of our psychology course, we have an almost total sample of first year psychology students at our university. In addition, we have data from more than a third of our third year psychology students. It is highly unlikely that other forms of recruiting for a qualitative study of student attitudes to research participation would be able to achieve this.

Thus, we can have some confidence that we have captured the views of the full range of UG psychology students at our university.

However, the use of exam answers as a data source is also a major limitation of this study. While the majority of students provided information on what they had learned through research participation, two students stated they had not learned anything, and a few more described their research participation experience in explicitly negative terms. The answers provided by some students may have been shaped by their perceptions of the types of responses that might earn “good marks”. We cannot know how many more students may have self-censored their negative views in an effort to please their examiners and/or maximise their exam marks.

Keeping in mind both this strength and limitation of using exam data, there are several conclusions we can draw from this research. On the surface, it certainly seems that most students appeared to have valued the opportunity to participate in research and reflected on their research participation. Further, there was evidence of consolidation of learning and the ability to recognise in others’ research the concepts taught in research methods classes. These findings are generally consistent with previous research indicating that at least some students perceive educational benefit in research participation (Davis & Fernald, 1975; Landrum & Chastain, 1995; Leak, 1981; Moreland 1999; Trafimow et al., 2006). However, the extent to which both participating in research and reflecting on their experiences was prompted by the knowledge of the exam question is unknown. It is possible that both participation rates and critical reflection may be lower when the “incentive” of knowing that the participation experience is examinable is removed. This is an area for future research.

In revisiting the risk-benefit ratio, it appears that research participation does offer educational benefits for at least some students. In our research, as in previous research (e.g., Coulter, 1986), we found that not all students viewed research participation as a positive experience or educational. If *some* students are experiencing educational benefits from research participation, can we build on this so that *all* students gain benefits, or so that students gain *increased* educational benefits? Previous researchers have suggested increasing the amount and type of feedback provided to student participants, both in the form of debriefing directly after each participation experience (Coulter, 1986; Davis & Fernald, 1975; Sieber, 1999) and later feedback (Dalziel, 1996). These activities may increase the pedagogic value associated with research participation.

Further gains may be achieved through tighter integration of research participation and the teaching of research methods. Dalziel (1996) suggested this could be achieved through the linking of discussions of theories, methods and methodological problems to available research projects. We recommend that teaching staff actively encourage students to participate in research as a way of supplementing the teaching and learning activities that are part of the standard research methods curricula. Active engagement as a research participant may promote deep learning opportunities, particularly where students are encouraged to actively reflect on their research participation experiences. Ideally, lecturers will be able to direct students towards a diversity of research participation opportunities and embed regular and ongoing discussion of these projects within the research methods curricula. Developing a culture of providing debriefing and feedback can be further enhanced through inviting researchers to present on both the research process and findings during research methods classes.

In summary, the research presented in this chapter aimed to examine psychology students' views of voluntary participation in research. The major theme to emerge was that participating in research provides increased insight into the research process. We propose that this educational gain may be further enhanced through providing greater integration of research participation and the teaching of research methods. Participating in psychological research offers benefits to both psychology student participants and researchers. We will leave the final word on this to one of the research participants in this study, who said:

By participating in research projects I was able to see and experience some of the hard work that goes into creating and carrying out a research project. In conclusion I have learnt that it is beneficial to both the researchers and myself to participate in research projects.

References

- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*, 77-101. doi:10.1191/1478088706qp063oa
- Britton, B. K., Richardson, D., Smith, S. S., & Hamilton, T. (1983). Ethical aspects of participating in psychological experiments: Effects of anonymity on evaluation, and complaints of distressed subjects. *Teaching of Psychology, 10*, 146-149. doi:10.1207/s15328023top1003_6
- Brody, J. L., Gluck, J. P., & Aragon, A. S. (2000). Participants' understanding of the process of psychological research: Debriefing. *Ethics & Behavior, 10*, 13-25. doi:10.1207/S15327019EB1001_2
- Clark, E., & McCann, T. (2005). Researching students: An ethical dilemma. *Nurse Researcher, 12*(3), 42-51.
- Coren, S. (1987). The psychology student subject pool: Student responses and attitudes. *Canadian Psychology, 28*, 360-363. doi:10.1037/h0084523
- Coulson, M. (1999). In at the deep end or still testing the water? Subject pools and required participation in UK psychology departments. *Psychology Teaching Review, 8*(1), 19-25.
- Coulter, X. (1986). Academic value of research participation by undergraduates. *American Psychologist, 41*, 317. doi:10.1037/0003-066X.41.3.317.a
- Dalziel, J. R. (1996). Students as research subjects: Ethical and educational issues. *Australian Psychologist, 31*, 119-123. doi:10.1080/00050069608260190
- Daugherty, T. K., & Lawrence, J.W. (1996). Short-term effects of research participation on college men. *Journal of Psychology, 130*, 71-77.

- Davis, J. R., & Fernald, P. S. (1975). Laboratory experience versus subject pool. *American Psychologist*, *30*, 523-524. doi:10.1037/h0076758
- Diamond, M. R., & Reidpath, D. D. (1992). Psychology ethics down under: A survey of student subject pools in Australia. *Ethics and Behavior*, *2*, 101-108. doi:10.1207/s15327019eb0202_3
- Elliott, L. J., Rice, S., Trafimow, D., Madson, L., & Hipshur, M. F. (2010). Research participation versus classroom lecture: A comparison of student learning. *Teaching of Psychology*, *37*, 129-131. doi:10.1080/00986281003626862
- Ferguson, L. M., Myrick, F., & Yonge, O. (2006). Ethically involving students in faculty research. *Nurse Education in Practice*, *6*, 397-403. doi:10.1016/j.nepr.2006.07.016
- Ferguson, L. M., Yonge, O., & Myrick, F. (2004). Students' involvement in faculty research: Ethical and methodological issues. *International Journal of Qualitative Methods*, *3*(4), 1-14. Retrieved from <http://ejournals.library.ualberta.ca/index.php/IJQM/article/view/4382>
- Fisher, C. B., & Fryberg, D. (1994). Participant partners: College students weigh the costs and benefits of deceptive research. *American Psychologist*, *49*, 417-427. doi:10.1037/0003-066X.49.5.417
- Flagel, D. C., Best, L. A., & Hunter, A. C. (2007). Perceptions of stress among students participating in psychology research: A Canadian survey. *Journal of Empirical Research on Human Research Ethics*, *2*(3), 61-67. doi:10.1525/jer.2007.2.3.61
- Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences*, *33*, 61-135. doi:10.1017/S0140525X0999152X

- Higbee, K., Millard, R., & Folkman, J. (1982). Social psychology research during the 1970s: Predominance of experimentation and college students. *Personality and Social Psychology Bulletin*, *8*, 180-183. doi:10.1177/014616728281028
- Kimble, G. A. (1987). The scientific value of undergraduate research participation. *American Psychologist*, *42*, 267-268. doi:10.1037/0003-066X.42.3.267.b
- King, D. J. (1970). The subject pool. *American Psychologist*, *25*, 1179-1181. doi:10.1037/h0037911
- Korn, J. H. (1999). Students' roles, rights, and responsibilities as research participants. In M. E. Ware & C. L. Brewer (Eds.), *Handbook for teaching statistics and research methods* (2nd ed., pp. 113-118). Mahwah, NJ: Lawrence Erlbaum.
- Landrum, R. E., & Chastain, G. D. (1995). Experiment spot-checks: A method for assessing the educational value of undergraduate participation in research. *IRB: Ethics and Human Research*, *17*(4), 4-6. Retrieved from <http://www.jstor.org/stable/3564152>
- Landrum, R. E., & Chastain, G. D. (1999). Subject pool policies in undergraduate-only departments: Results from a nationwide survey. In G. D. Chastain & R. E. Landrum (Eds.), *Protecting human subjects: Departmental subject pools and institutional review boards* (pp. 25-43). Washington, DC: American Psychological Association.
- Leak, G. K. (1981). Student perception of coercion and value from participation in psychological research. *Teaching of Psychology*, *8*, 147-149. doi:10.1207/s15328023top0803_4

- Lindsay, R. C., & Holden, R. R. (1987). The introductory psychology subject pool in Canadian universities. *Canadian Psychology/Psychologie Canadienne*, 28, 45-52. doi:10.1037/h0079868
- McNemar, Q. (1946). Opinion-attitude methodology. *Psychological Bulletin*, 43, 289-374. doi:10.1037/h0060985
- Midzinski, A. L. (2010). *Students' perceptions of coercion in research* (Master's thesis, Utah State University). Retrieved from <http://digitalcommons.usu.edu/etd/583/>
- Miller, W. E., & Kreiner, D. S. (2008). Student perception of coercion to participate in psychological research. *North American Journal of Psychology*, 10, 53-64.
- Moreland, R. L. (1999). Evaluating students' research experiences via credit slips. In G. D. Chastain & R. E. Landrum (Eds.), *Protecting human subjects: Departmental subject pools and institutional review boards* (pp. 87-108). Washington, DC: American Psychological Association.
- National Health and Medical Research Council, Australian Research Council, & Australian Vice-Chancellors' Committee. (2007). *National statement on ethical conduct in human research*. Canberra, Australia: Author. Retrieved from <http://www.nhmrc.gov.au/publications/synopses/e72syn.htm>
- Nimmer, J. G., & Handelsman, M. M. (1992). Effects of subject pool policy on student attitudes toward psychology and psychological research. *Teaching of Psychology*, 19, 141-144. doi:10.1207/s15328023top1903_2
- Norenzayan, A., & Heine, S. J. (2005) Psychological universals: What are they and how can we know? *Psychological Bulletin*, 131, 763-784. doi:10.1037/0033-2909.131.5.763

- Richardson, D. R., Pegalis, L., & Britton, B. (1992). A technique for enhancing the value of research participation. *Contemporary Social Psychology, 16*(1-2), 11-13.
- Sears, D. O. (1986). College sophomores in the laboratory: Influences of a narrow data base on social psychology's view of human nature. *Journal of Personality and Social Psychology, 51*, 515-530. doi:10.1037/0022-3514.51.3.515
- Shi, L. (2006). Students as research participants or as learners? *Journal of Academic Ethics, 4*, 205-220. doi:10.1007/s10805-006-9028-y
- Sieber, J. E. (1999). What makes a subject pool (un)ethical. In G. D. Chastain & R. E. Landrum (Eds.), *Protecting human subjects: Departmental subject pools and institutional review boards* (pp. 43-68). Washington, DC: American Psychological Association.
- Sieber, J. E., & Saks, M. J. (1989). A census of subject pool characteristics and policies. *American Psychologist, 44*, 1053-1061. doi:10.1037/0003-066X.44.7.1053
- Sullivan, M. P., & Lashley, S. L. (2009). Developing informed research participants in an introductory psychology participant pool. *Teaching of Psychology, 36*, 24 – 28. doi:10.1080/00986280802529269
- Tabachnick, B. G., Keith-Spiegel, P., & Pope, K. S. (1991). Ethics of teaching: Beliefs and behaviors of psychologists as educators. *American Psychologist, 46*, 506-515. doi:10.1037/0003-066X.46.5.506
- Trafimow, D., Madson, L., & Gwizdowski, L. (2006). Introductory psychology students' perceptions of alternatives to research participation. *Teaching of Psychology, 33*, 247-249. doi:10.1207/s15328023top3304_7

Wintre, M. G., North, C., & Sugar, L. A. (2001). Psychologists' response to criticisms about research based on undergraduate participants: A developmental perspective. *Canadian Psychology/Psychologie Canadienne*, 42, 216-225.
doi:10.1037/h0086893

**Paper 2: A Brief Measure of Student Perceptions of the Educational Value of
Research Participation**

Roberts, L. D., & **Allen, P. J.** (2013). A brief measure of student perceptions of the educational value of research participation. *Australian Journal of Psychology*, 65, 22-29. doi:10.1111/ajpy.12007

This is an electronic version of an article published in: *Australian Journal of Psychology*, 2013; 65, 22-29. doi:10.1111/ajpy.12007

In compliance with the Australian Psychological Society's terms and conditions of use, this is the post-print version of the abovementioned work. Readers wishing to cite this paper are encouraged to source the final published version, available at: <http://onlinelibrary.wiley.com/doi/10.1111/ajpy.12007/abstract>

**A Brief Measure of Student Perceptions of the Educational Value of Research
Participation**

Lynne D. Roberts and Peter J. Allen

**School of Psychology and Speech Pathology, Curtin Health Innovation Research
Institute, Curtin University**

Address for correspondence:

Lynne Roberts (PhD)

Senior Lecturer, School of Psychology and Speech Pathology

Curtin Health Innovation Research Institute, Curtin University

GPO Box U1987

Perth, WA, 6845, Australia

Email: Lynne.Roberts@curtin.edu.au

Tel: +61 8 9266 7183

Fax: +61 8 9266 2464

Peter Allen: p.allen@curtin.edu.au

Abstract

Despite the continued reliance on undergraduate students as research participants there is an absence of valid, reliable measures of student perceptions of educational gains from research participation. In this paper we present two studies outlining the development and initial validation of a new measure, the *Student Perceptions of the Educational Value of Research Participation Scale*. In Study One a pool of 28 items was developed from previous qualitative research and administered to a convenience sample of 68 Australian university student volunteers. Following Principal Axis Factoring, a seven item uni-dimensional scale with good internal reliability ($\alpha = .82$) was developed and validated against an existing measure of reactions to research participation. In Study Two, 104 members of a second year undergraduate psychology participant pool completed the measure. Confirmatory factor analysis supported a higher order two factor model (overall $\alpha = .82$). Across both volunteer and participant pool samples, the educational value of research participation was rated more highly than the costs of research participation (emotional reactions and drawbacks of participating), indicating a positive cost-benefit ratio of research participation. This brief, internally reliable measure can be used in assessing students' perceptions of educational gain in both individual research projects and across research projects.

A Brief Measure of Student Perceptions of the Educational Value of Research Participation

The majority of published psychological research is based on the use of undergraduate student participants (Higbee, Millard, & Folkman, 1982; Korn, 1999; Sieber & Saks, 1989; Wintre, North, & Sugar, 2001), who are easily accessible to university-based researchers (Diamond & Reidpath, 1992; Sieber & Saks, 1989). Many universities utilise participant pools with mandated research participation for first year students. Penalties may be applied for non-participation (Diamond & Reidpath, 1992). Even where participation is not mandated, incentives such as extra credit may be offered as an inducement to participate in research (Sieber & Saks, 1989).

This reliance on undergraduate students as research participants has been questioned on both research and ethical grounds. From a research perspective, the generalisability of results based on undergraduate students has been questioned, with the potential for systematic biases in research findings identified as a particular concern (Henrich, Heine, & Norenzayan, 2010; Norenzayan & Heine, 2005; Sears, 1986). From an ethical perspective, concerns relate to dual relationships (Clark & McCann, 2005; Ferguson, Myrick, & Yonge, 2006; Ferguson, Yonge, & Myrick, 2004; Shi, 2006) and the perceptions of coercion associated with mandated and incentivised research participation (Miller & Kreiner, 2008). However, others (e.g., Dalziel, 1996) have argued that mandated and incentivised research participation for students is no more coercive than other course requirements such as exams and essays. Instead, Dalziel argued that the question is whether research participation is educationally justifiable.

The educational benefits suggested to result from the experiential learning involved in research participation include exposure to a variety of research methods and research processes that may develop students' research capabilities (Dalziel, 1996; Moreland, 1999). As such, research participation may contribute to both meeting the graduate attribute of

Research Methods in Psychology (see Morris, Cranney, Jeong, & Mellish, this issue) and the development of psychological literacy, important learning outcomes for undergraduate psychology students (Cranney & Dunn, 2011; Morris et al., this issue). Qualitative and survey research findings to date indicate that some, but not all, students see research participation as having some educational value (e.g., Darling, Goedert, Ceynar, Shore, & Anderson, 2007), with benefits including learning about psychology, understanding research and ethical processes, and increasing interest in psychology (Landrum & Chastain, 1995; Moreland 1999; Rosell, Beck, Luther, Goedert, Shore, & Anderson, 2005; Trafimow, Madson, & Gwizdowski, 2006). Experimental research has provided limited support for the educational value of participating in research. Gil-Gómez de Liaño, León, and Pascual-Ezama (2012) reported that students who participated in research (regardless of whether or not they engaged in activities designed to enhance their understanding of the experiments they participated in) scored higher on a research methods exam than those who did not. However, students self-selected into the ‘no participation’ condition, thus systematically differing in motivation from students in the research participation conditions. Elliott, Rice, Trafimow, Madson, and Hipshur (2010) randomly assigned students to a lecture or participation in an experiment followed by debriefing. While there was a main effect for knowledge gained from pre-test to post-test, there was no significant difference across conditions, suggesting no advantage of research participation over lecture attendance.

Ethical research balances potential benefits from research against potential risks to research participants and others. Student research participation provides clear benefits to researchers through providing access to research participants, with flow on benefits to society through the findings stemming from the research. However, the benefits to students in terms of educational gains are seldom assessed (Landrum & Chastain, 1999) and have not been adequately measured (typically relying on a very limited number of survey items), inhibiting

the ability to conduct a full cost-benefit analysis of student research participation. In addition to research measuring actual education gains from research participation, there is a need for a psychometrically sound measure of student perceptions of education gains from research participation that can be used across studies and contexts.

In a previous qualitative study (Roberts & Allen, 2012), we asked 195 undergraduate students, “You’ve been invited to participate in a number of research projects this semester; what have you learned from this experience?” Their responses were content analysed, and the key theme that emerged was that through participating in others’ research, students gained insight into the research process, enhancing their knowledge about the conduct of psychological research.

Within this overarching theme of insight into the research process, eight sub-themes were identified. The first six themes comprised insight into: the range of topics and issues that were studied by psychologists and within the university; how ‘real’ research was conducted; the role of the researcher; the opportunity to view participation from the participant perspective; process issues in conducting research; and specific information on research methodologies (predominantly questionnaires). The seventh sub-theme was participation as complementing the research methods taught in class and the eighth was that the insights made could be applied in their own future research (Roberts & Allen, 2012).

In this article we draw on the major theme and subthemes identified to develop and begin the validation of a new scale, the *Student Perceptions of the Educational Value of Research Participation Scale (SPEVRPS)*, which measures students’ perceived educational gains from research participation. In a search of the literature we could locate no measure of student perceptions of the educational value of research that could be used across disciplines and types of research. Given the concerns raised over whether student research participation

is educationally and ethically justifiable, such a measure is required to begin to examine the cost-benefit ratio of student research participation.

Study One

The primary aim of Study One was to develop a brief measure of student perceptions of the educational value of research participation. In addition, this study begins the validation of the new measure, with the convergent and divergent validity of the *SPEVRPS* assessed in relation to the subscales of the *Reactions to Research Participation Questionnaire-Revised* (*RRPQ-R*; Newman, Willard, Sinclair, & Kaloupek, 2001), a questionnaire designed to measure participants' evaluation of research participation, but that does not explicitly measure perceived educational gains. It was hypothesised that correlations between the new measure and the *RRPQ-R* would be weak and positive with the 'Participation' and 'Personal Benefits' subscales, weak and negative with the 'Emotional Reactions' and 'Perceived Drawbacks' subscales, and moderate and positive with the 'Global Evaluation' subscale.

As a further form of validation, the relationship between scores on the new measure and amount of research participation was examined. At the time of this first study (2010) there was no participant pool in our university, and volunteers were used in all research. It is likely that only students who perceive some sort of benefit from research participation would volunteer to participate in multiple research projects. Thus, it was hypothesised that there would be a positive correlation between scores on the new measure and the number of studies participated in over the semester.

The secondary aim of this study was to compare students' perceptions of the educational value of research against their perceptions of the costs of the research. It was hypothesised that the perceived educational benefits of research (measured by the newly developed *SPEVRPS*) would be greater than the perceived costs of participation (measured by the 'Emotional Reactions' and 'Drawbacks' subscales of the *RRPQ-R*).

Method

A cross-sectional correlational design was utilised with data collected using an online survey.

Participants.

Participants were a convenience sample of 68 Australian undergraduate psychology students. Reflecting the gender bias in the psychology student population, 83.8% of the research participants were female (male 13.2%, unstated 2.9%). Participants ranged in age from 17 to 51 ($M = 24.7$, $SD = 8.34$). Almost half (45.6%) were first year students, with second year (19.1%), third year (11.8%) and fourth year (20.6%) undergraduate students also represented. All participants had taken part in at least one previous study during the semester (Range = 1-12; Median = 3). The majority (98.5%) had participated in questionnaire or survey research, 23.5% had participated in experimental research and 5.9% had participated in a qualitative study.

Measures.

An online questionnaire was constructed containing 28 items designed to measure student perceptions of the educational value of research participation, the *RRPQ-R*, single item measures of demographics (age, gender, year of study) and the number and type of research projects participated in.

Perceptions of the educational value of research. Based on our prior qualitative analysis (Roberts & Allen, 2012) we developed a pool of 28 items designed to provide measures of seven of the eight sub-themes of insight into the research process¹. Items were reviewed to ensure they represented views expressed by participants in the original qualitative study. Each was rated on a scale ranging from (1) *strongly disagree* to (5) *strongly agree*.

¹ We developed four additional items linked to the final sub-theme, which was about questionnaire design. However, these were excluded from subsequent analyses, as our aim was to develop a measure that could be used across all types of research.

RRPQ-R (Newman et al., 2001). The *RRPQ-R* is a 23-item scale designed to measure ethical constructs associated with research participation. The measure comprises five subscales named 'Participation' (4 items; example item "I like the idea that I contributed to science"), 'Personal Benefits' (4 items; example item "I gained insight about my experiences through research participation"), 'Emotional Reactions' (4 items; example item "The research made me think about things I didn't want to think about"), 'Perceived Drawbacks' (6 items; example item "I found the questions too personal"), and 'Global Evaluation' (5 items; example item "I was treated with respect and dignity"). One item was reworded from "I understood the consent form" to "I understood the information sheet/consent form". In the current study, each item was rated on a scale from (1) *strongly disagree* to (5) *strongly agree*.

Previous exploratory and confirmatory factor analyses support the five factor structure of the *RRPQ-R* (Newman et al., 2001). When used in previous research, the internal reliabilities of the subscales have mostly fallen within acceptable ranges ($\alpha = .53$ to $.87$; DePrince & Chu, 2008; Edwards, Kearns, Calhoun, & Gidycz, 2009; Newman et al., 2001).

Procedure.

Prior to commencing the research, ethics approval was obtained from Curtin University Human Research Ethics Committee. To prevent perceptions of coercion that could arise from our dual roles as lecturers and researchers, data collection was online, anonymous and completed outside of class time. Recruitment for the research commenced late in Semester 2, 2010, through advertisements on student learning management system sites and announcements during lectures. Participation in the research was voluntary and no incentive for participation was offered. Interested students were provided with a link to an online participant information sheet, and upon consenting were redirected to an online questionnaire.

Of 82 survey responses, 11 were deleted because the *SPEVRPS* items had not been completed and three cases were removed that did not meet the inclusion criteria. Across the survey there were 24 missing data points (0.06% missing data). A missing values analysis indicated that these data were missing completely at random, Little's MCAR test $\chi^2(635, N = 68) = 652.66, p = .276$. Missing data points on scale measures were replaced using Expectation-Maximization.

Results

Principal axis factoring with varimax rotation was used to explore the factor structure of the 28 newly created student perceptions of the educational value of research participation items. Nine factors were extracted with Eigenvalues greater than one. However, the unrotated factor solution (with most items loading more strongly on the first factor than other factors) and scree plot (steep curve between first factor with an eigenvalue of 9.5 and all other factors with eigenvalues between 2.2 and 1.0) suggested only one main factor underlying the items. In order to develop a short measure with strong content validity, the highest loading item for each of the seven sub-themes (see Table 1) were selected for a second factor analysis. Principal axis factoring extracted one factor with an Eigenvalue over 1 with all item loadings above .4, confirming the uni-dimensionality of the measure. The items and factor loadings are presented in Table 1. The Cronbach's alpha for this factor was .82, suggesting good internal reliability in this sample.

<insert Table 1 about here>

For each respondent, the mean of the seven items was calculated to provide a scale score. This new measure is called the *Student Perceptions of the Educational Value of Research Participation Scale (SPEVRPS)*. Descriptive statistics for the new measure and the *RRPQ-R* are presented in Table 2.

<insert Table 2 about here>

The relationships between the *SPEVRPS* and the subscales of the *RRPQ-R* were examined to assess convergent and divergent validity through a series of bivariate correlations ($N = 68$). As hypothesised, there were significant positive correlations between the *SPEVRPS* and ‘Participation’ ($r = .27, p = .024$), ‘Personal Benefits’ ($r = .33, p = .006$) and ‘Global Evaluation’ ($r = .44, p < .001$) subscales of the *RRPQ-R*. There was a significant negative correlation between the *SPEVRPS* and the ‘Perceived Drawbacks’ subscale ($r = -.25, p = .041$) and a negative, but non-significant, correlation with the ‘Emotional Reactions’ subscale ($r = -.14, p = .243$) of the *RRPQ-R*.

The *SPEVRPS* was also examined in relation to demographic and participation variables ($N = 68$). As hypothesised, there was a positive correlation between the *SPEVRPS* and the number of research projects participated in over semester ($r_s = .32, p = .007$). There was no significant relationship with age ($r_s = .08, p = .440$) or gender, $t(64) = -0.97, p = .337$.

To compare students’ perceptions of the educational value of research against their perceptions of the costs of the research, two paired samples *t*-tests were conducted, comparing scores on the *SPEVRPS* to scores on the ‘Emotional Reactions’ and ‘Drawbacks’ subscales of the *RRPQ-R*. As hypothesised, scores on the *SPEVRPS* ($M = 4.01, SD = .49$) were significantly higher than scores on the ‘Emotional Reactions’ subscale ($M = 1.77, SD = .69$), $t(67) = 20.40, p < .001, d = 3.80$, and the ‘Perceived Drawbacks’ subscale ($M = 2.09, SD = .55$), $t(67) = 19.21, p < .001, d = 3.69$, of the *RRPQ-R*.

Discussion

In Study One we have presented the development and initial validation of a new brief measure of student perceptions of the education value of research, the *SPEVRPS*. This measure has good content validity, covering seven subthemes of ‘insight into the research process’ that emerged as the key educational outcome of research participation in previous qualitative research (Roberts & Allen, 2012). Exploratory factor analysis and internal

reliability testing indicated the measure is uni-dimensional and has good internal consistency. The *SPEVRPS* correlated with subscales of the *RRPQ-R* in the expected directions, supporting the convergent and divergent validity of the measure. In addition, scores on the *SPEVRPS* were positively correlated with the number of research studies participated in over the semester, and students rated the perceived benefits of participating in research higher than the perceived costs.

Further validation of the measure is required. The sample for this study comprised students who volunteered to participate in research. It will be important to examine whether the psychometric properties of the measure hold when the sample comprises students who are members of a participant pool.

Study Two

The aim of Study Two was to continue the validation of the *SPEVRPS* on a sample of undergraduate psychology students recruited through a participant pool. The introduction of an undergraduate participant pool to our university in 2012 provided an ideal opportunity to collect comparative data to Study One. This study examines the factor structure and the test-retest reliability of the measure within a participant pool sample. It further examines the relationship between the *SPEVRPS* and the *RRPQ-R* to determine if the benefits of participating in research continue to outweigh the costs of participating in research for members of a participant pool.

Method

A cross-sectional correlational design was utilised with data collected using an online survey.

Participants.

All students in the second year undergraduate psychology participant pool ($N = 144$) had the option of participating in this study. Of these, 104 (73.5% female) completed the

survey, providing a response rate of 72%. Participants ranged in age from 18 to 47 ($M = 20.7$, $SD = 4.04$). All participants had taken part in at least one previous research study within the university that semester (Range = 1-9; Median = 5). All had participated in questionnaire research, 39.8% had participated in experimental research and 6.1% had participated in qualitative research.

Measures.

An online questionnaire was constructed containing the new seven item *SPEVRPS*, the *RRPQ-R* and demographic questions.

Procedure.

Ethics approval was obtained from the Curtin University Human Research Ethics Committee. Recruitment for the research commenced in Semester 1, 2012, through the second year participant pool. Interested students were provided with a link to an online participant information sheet, which linked them to an online questionnaire. Students completing the survey were assigned credits towards their research participation requirement. Students not participating in this research could participate in other research studies or complete an alternative activity. At the end of the questionnaire, participants were invited to indicate their interest in completing a second survey two weeks later, containing the *SPEVRPS* only. Twenty-two participants completed the second survey between 13 and 17 days after completion of the initial survey.

There were three missing data points across the scales, and these were replaced using mean substitution. Missing data points on demographic questions (gender = 6, age = 8) were not replaced.

Results

Confirmatory factor analysis of the *SPEVRPS* was conducted using EQS (version 6.2). A one factor model, as suggested by Study One, was tested against the recommended

cut-offs for goodness of fit of four fit indices: the Satorra-Bentler Chi Square divided by degrees of freedom, the Comparative Fit Index (CFI), the Non-Normed Fit Index (NNFI) and the Root Mean Square Error of Approximation (RMSEA). The fit statistics suggested the fit of the one factor model was less than optimal (see Table 3). Following exploratory factor analysis, which suggested a two factor model may provide a better fit to the data, three further models were tested using confirmatory factor analysis: a correlated two-factor model, an uncorrelated two factor model and a higher order model. Fit indices for each of the models are presented in Table 3. The higher order model (see Figure 1) is preferred to the other models because of superior fit indices. The first factor relates to the practice of research and the second factor to knowledge of research. The full measure has good internal reliability, and as such is suitable for use as a single measure. Descriptive statistics and reliability coefficients for the *SPEVRPS* and the *RRPQ-R* in this sample are presented in Table 2.

<insert Table 3 about here>

<insert Figure 1 about here>

The relationships between the *SPEVRPS* and the subscales of the *RRPQ-R* were examined to determine convergent and divergent validity through a series of bivariate correlations ($N = 104$). As hypothesised, there were significant positive correlations between the *SPEVRPS* and 'Participation' ($r = .47, p < .001$), 'Personal Benefits' ($r = .60, p < .001$) and the 'Global Evaluation' ($r = .23, p = .03$) subscales of the *RRPQ-R*. There was a significant negative correlation between the *SPEVRPS* and the 'Perceived Drawbacks' subscale ($r = -.42, p < .001$) and an unexpected positive correlation with the 'Emotional Reactions' subscale ($r = .23, p = .02$) of the *RRPQ-R*. There was no significant relationship between the *SPEVRPS* and the number of research projects participated in over semester ($r = .04, p = .71$), age ($r_s = -.24, p = .82$) or gender, $t(96) = -1.97, p = .06$.

To examine test-retest reliability, the Time 1 and Time 2 scores were correlated for the 22 participants who completed both surveys. A moderate significant correlation was found ($r = .47, p = .03$).

To compare students' perceptions of the educational value of research against their perceptions of the costs of the research, two paired samples *t*-tests were conducted, comparing scores on the *SPEVRPS* to scores on the 'Emotional Reactions' and 'Drawbacks' subscales of the *RRPQ-R*. Scores on the *SPEVRPS* ($M = 3.90, SD = .48$) were significantly higher than scores on the 'Emotional Reactions' subscale ($M = 2.01, SD = .73$), $t(103) = 24.81, p < .001, d = 3.12$, and the 'Perceived Drawbacks' subscale ($M = 2.54, SD = .69$), $t(103) = 14.01, p < .001, d = 2.32$, of the *RRPQ-R*.

Comparison across Samples.

Independent samples *t*-tests were conducted to compare the scores on the *SPEVRPS* and *RRPQ-R* across the volunteer and participant pool samples. The results are summarised in Table 2. A consistent pattern of lower scores on measures of the benefits of research participation and higher scores on the disadvantages of research participation in the participant pool sample in comparison to the volunteer sample emerged, with small to large effect sizes.

Discussion

In Study Two, using a sample of students recruited through an undergraduate psychology participant pool, the findings from Study One were largely replicated. Confirmatory Factor Analysis indicated a higher order model best represented the measure. The internal reliability of the overall scale remained high. The test-retest reliability, conducted on a small sample, was lower than desirable and requires further testing on a larger sample. The *SPEVRPS* mostly correlated with subscales of the *RRPQ-R* in the expected directions, supporting the convergent and divergent validity of the measure. The positive

correlation between the *SPEVRPS* and 'Emotional Reactions' subscale was unexpected, and further research is required to see if this finding is replicated. Students rated the perceived benefits of participating in research more highly than the perceived costs of participating in research.

Overall Discussion

The *SPEVRPS* is a brief, reliable measure of student perceptions of the educational value of research participation. Combined with the promising psychometric properties, the brevity of the *SPEVRPS* (seven items) means that it may be suitable for use both within individual research projects and across research projects. It could be included at the end of individual research projects to provide an indication of the perceived educational value of the particular project. Alternatively, the *SPEVRPS* is suitable for assessing students' perceptions of the educational value of participation across multiple studies, and could be useful to researchers seeking to develop an evidence base to support the integration of participant pools in undergraduate psychology courses that is comparable to the evidence bases that have been developed around other pedagogic tools and techniques common to the modern psychology classroom (see, for e.g., Dunn, Saville, Baker, & Marek, this issue).

Identification of how students perceive the educational value of research participation provides an important first step in evaluating the possible benefits of research participation. However, this should not be seen as an end in itself. Previous research has suggested that not all students view research participation as positive or educational (Coulter, 1986; Roberts & Allen, 2012). Increasing the pedagogic value of student research participation for all students may require a closer integration of research participation and the teaching of research methods (Dalziel, 1996) through the provision of debriefing directly after each participation experience (Coulter, 1986; Sharpe & Faye, 2009; Sieber, 1999), later feedback (Dalziel, 1996) and class assignments relating to the participation experience (Moyer & Franklin,

2011). Increasing the educational value of research participation will contribute to developing the psychological literacy of undergraduate students, through enhancing knowledge and understanding of scientific and ethical research practices.

The *SPEVRPS* can be used on its own or in conjunction with the *RRPQ-R* to provide a more comprehensive measure of the costs and benefits of research participation. The *SPEVRPS* provides researchers with the opportunity to revisit the key ethical question of whether potential benefits from research outweigh potential risks to research participants. The findings from these initial studies suggest that students rate the perceived benefits of participating in research more highly than the perceived costs of participating in research. However, the ratio of benefits to costs varied across samples, with the participant pool sample perceiving less benefits and greater costs than the volunteer sample. Future research could examine whether this ratio changes across participation in research of differing levels of sensitivity, requiring differing commitment of time or effort, or using different methodologies (e.g., experimental versus survey research).

In summary, this paper has presented the development and initial validation of a measure of the perceived educational value of research participation, the *SPEVRPS*. This is a brief, internally reliable measure that can be used in assessing students' perceptions of educational gain in both individual research projects and across research projects.

References

- Clark, E., & McCann, T. (2005). Researching students: An ethical dilemma. *Nurse Researcher, 12*(3), 42-51.
- Coulter, X. (1986). Academic value of research participation by undergraduates. *American Psychologist, 41*, 317. doi:10.1037//0003-066X.41.3.317.a
- Cranney, J., & Dunn, D. (Eds.) (2011). *The psychologically literate citizen: Foundations and global perspectives*. New York, NY: Oxford University Press.
- Dalziel, J. R. (1996). Students as research subjects: Ethical and educational issues. *Australian Psychologist, 31*, 119-123. doi:10.1080/00050069608260190
- Darling, J., Goedert, K., Ceynar, M., Shore, W., & Anderson, D. (2007). Learning about the means to an end: What US introductory psychology students report about experimental participation. *Psychology Learning and Teaching, 6*, 91-97. doi:10.2304/plat.2007.6.2.91
- DePrince, A. P., & Chu, A. (2008). Perceived benefits in trauma research: Examining methodological and individual difference factors in responses to research participation. *Journal of Empirical Research on Human Research, 3*, 35-47. doi:10.1525/jer.2008.3.1.35
- Diamond, M. R., & Reidpath, D. D. (1992). Psychology ethics down under: A survey of student subject pools in Australia. *Ethics and Behavior, 2*, 101-108. doi:10.1207/s15327019eb0202_3
- Edwards, K. M., Kearns, M. C., Calhoun, K. S., & Gidycz, C. A. (2009). College women's reactions to sexual assault research participation: Is it distressing? *Psychology of Women Quarterly, 33*, 225-234. doi:10.1111/j.1471-6402.2009.01492.x

- Elliott, L. J., Rice, S., Trafimow, D., Madson, L., & Hipshur, M. F. (2010). Research participation versus classroom lecture: A comparison of student learning. *Teaching of Psychology, 37*, 129-131. doi:10.1080/00986281003626862
- Ferguson, L. M., Myrick, F., & Yonge, O. (2006). Ethically involving students in faculty research. *Nurse Education in Practice, 6*, 397-403. doi:10.1016/j.nepr.2006.07.016
- Ferguson, L. M., Yonge, O., & Myrick, F. (2004). Students' involvement in faculty research: Ethical and methodological issues. *International Journal of Qualitative Methods, 3*(4), 1-14.
- Gil-Gómez de Liaño¹, B., León, O. F., & Pascual-Ezama, D. (2012). Research participation improves student's exam performance. *The Spanish Journal of Psychology, 15*, 544-550. doi:10.5209/rev_SJOP.2012.v15.n2.38865
- Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences, 33*, 61-135. doi:10.1017/S0140525X0999152X
- Higbee, K., Millard, R., & Folkman, J. (1982). Social psychology research during the 1970s: Predominance of experimentation and college students. *Personality and Social Psychology Bulletin, 8*, 180-183. doi:10.1177/014616728281028
- Korn, J. H. (1999). Students' roles, rights, and responsibilities as research participants. In M. E. Ware & C. L. Brewer (Eds.), *Handbook for teaching statistics and research methods* (2nd ed., pp. 113-118). Mahwah, NJ: Lawrence Erlbaum.
- Landrum, R. E., & Chastain, G. (1999). Subject pool policies in undergraduate-only departments: Results from a nationwide survey. In G. Chastain & R. E. Landrum (Eds.), *Protecting human subjects: Departmental subject pools and institutional review boards* (pp. 25-43). Washington, DC: American Psychological Association.
- Miller, W. E., & Kreiner, D. S. (2008). Student perception of coercion to participate in psychological research. *North American Journal of Psychology, 10*, 53-64.

- Moreland, R. L. (1999). Evaluating students' research experiences via credit slips. In G. Chastain & R. E. Landrum (Eds.), *Protecting human subjects: Departmental subject pools and institutional review boards* (pp. 87-108). Washington, DC: American Psychological Association.
- Moyer, A., & Franklin, N. (2011). Strengthening the educational value of undergraduate participation in research as part of a psychology department subject pool. *Journal of Empirical Research on Human Research Ethics*, 6, 75-82. doi:10.1525/jer.2011.6.1.75
- Newman, E., Willard, T., Sinclair, R., & Kaloupek, D. (2001). Empirically supported ethical research practice: The costs and benefits of research from the participants' view. *Accountability in Research*, 8, 309-329. doi:10.1080/08989620108573983
- Norenzayan, A., & Heine, S. J. (2005) Psychological universals: What are they and how can we know? *Psychological Bulletin*, 131, 763–84. doi:10.1037/0033-2909.131.5.763
- Roberts, L. D., & Allen, P. J. (2012). Student perspectives on the value of research participation. In S. McCarthy, K. L. Dickson, J. Cranney, A Trapp, & V. Karandashev (Eds.), *Teaching psychology around the world* (Volume 3, pp. 198-212). Newcastle upon Tyne, UK: Cambridge Scholars Publishing.
- Rosell, M. C., Beck, D. M., Luther, K. E., Goedert, K. M., Shore, W. J., & Anderson, D. D. (2005). The pedagogical value of experimental participation paired with course content. *Teaching of Psychology*, 32, 95-99. doi:10.1207/s15328023top3202_3
- Sears, D. O. (1986). College sophomores in the laboratory: Influences of a narrow data base on social psychology's view of human nature. *Journal of Personality and Social Psychology*, 51, 515-530. doi:10.1037/0022-3514.51.3.515
- Sharpe, D., & Faye, C. (2009). A second look at debriefing practices: Madness in our method? *Ethics & Behavior*, 19, 432-447. doi:10.1080/10508420903035455

- Shi, L. (2006). Students as research participants or as learners? *Journal of Academic Ethics*, 4, 205-220. doi:10.1007/s10805-006-9028-y
- Sieber, J. E. (1999). What makes a subject pool (un)ethical. In G. Chastain & R. E. Landrum (Eds.), *Protecting human subjects: Departmental subject pools and institutional review boards* (pp. 43-68). Washington, DC: American Psychological Association.
- Sieber, J. E., & Saks, M. J. (1989). A census of subject pool characteristics and policies. *American Psychologist*, 44, 1053-1061. doi:10.1037/0003-066X.44.7.1053
- Trafimow, D., Madson, L., & Gwizdowski, L. (2006). Introductory psychology students' perceptions of alternatives to research participation. *Teaching of Psychology*, 33, 247-249. doi:10.1207/s15328023top3304_7
- Wintre, M. G., North, C., & Sugar, L. A. (2001). Psychologists' response to criticisms about research based on undergraduate participants: A developmental perspective. *Canadian Psychology/Psychologie Canadienne*, 42, 216-225.

Table 1

Sub-Themes, Selected Items and Factor Loadings for the Seven Item Student Perceptions of the Educational Value of Research Participation Scale

Sub-theme	Item	Factor Loading
Complements teaching	I have been able to put what we have learned in class into context	.79
Process issues	I have seen how to put research methods concepts into practice	.78
Participant's perspective	I was able to see research from the perspective of a participant	.66
Conduct of 'real research'	I increased my knowledge of how research is conducted	.65
Role of researcher	I have learned how to behave as a researcher	.56
Own future research	I feel I will be able to put the knowledge gained into practice in future years	.55
Varieties of research	I increased my knowledge about the range of research conducted in my university	.44

Table 2

Descriptive Statistics and Reliability Coefficients for the SPEVRPS and RRPQ-R in Study One and Study Two

Measure	Study 1 (<i>N</i> = 68)			Study 2 (<i>N</i> = 104)			<i>d</i>
	α	Range	Mean (<i>SD</i>)	α	Range	Mean (<i>SD</i>)	
SPEVRPS	.82	2.43-5.00	4.01(.49)	.82	2.14-5.00	3.90(.48)	.23
Factor 1				.79	1.25-5.00	3.76(.58)	
Factor 2				.69	3.00-5.00	4.09(.50)	
RRPQ-R							
Participation	.68	3.00-5.00	4.26(.55)	.72	1.75-5.00	3.69(.72)**	.90
Personal Benefits	.82	1.75-4.75	3.55(.69)	.83	1.00-4.75	3.33(.72)	.31
Emotional Reactions	.92	1.00-4.00	1.77(.69)	.89	1.00-4.00	2.01(.73)*	-.34
Perceived Drawbacks	.76	1.00-3.50	2.09(.55)	.84	1.17-4.33	2.54(.69)**	-.73
Global Evaluation	.80	2.80-5.00	4.31(.49)	.83	3.00-5.00	4.13(.48)*	.37

Note. α = Cronbach's alpha. *d* = Cohen's *d* indexing the standardized difference between the relevant Study 1 and Study 2 means.

p* < .05. *p* < .001.

Table 3

Fit Indices for Confirmatory Factor Analysis Models (Robust Statistics)

Model	S-B χ^2/df	CFI	NNFI	RMSEA
Cut-off Criteria	$p > .05$	$\geq .85$	$\geq .85$	$\leq .06$
One Factor Model	$p < .01$	0.86	0.78	0.11
Correlated 2 Factor Model	$p = .30$	0.98	0.97	0.04
Uncorrelated 2 Factor Model	$p < .001$	0.78	0.67	0.14
Higher Order Model	$p = .78$	1.00	1.06	0.00

Note. S-B = Satorra-Bentler. CFI = Comparative Fit Index. NNFI = Non-Normed Fit Index. RMSEA = Root Mean Square Error of Approximation.

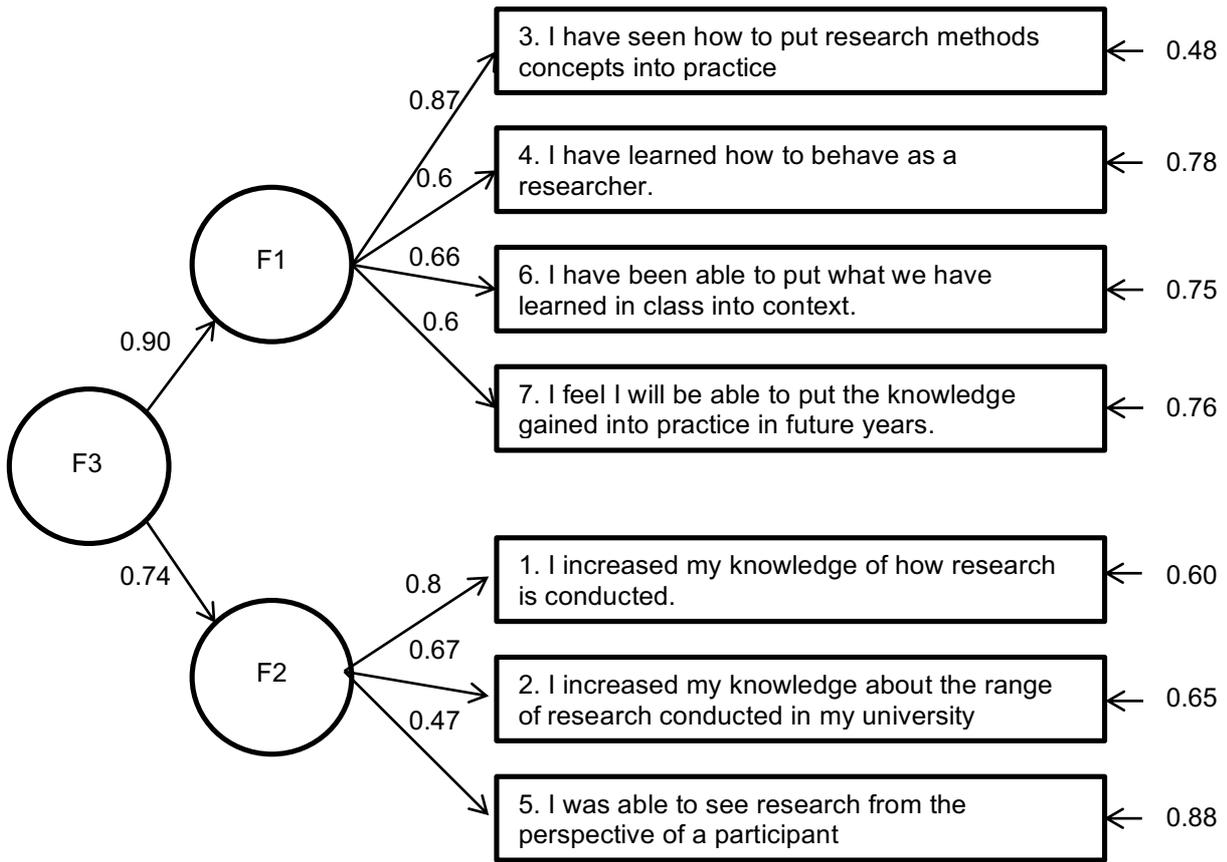


Figure 1. Higher order confirmatory factor model.

Paper 3: Active Learning in Research Methods Classes is Associated with Higher Knowledge and Confidence, though not Evaluations or Satisfaction

Allen, P. J., & Baughman, F. D. (2016). Active learning in research methods classes is associated with higher knowledge and confidence, though not evaluations or satisfaction. *Frontiers in Psychology*, 7, Article 279. doi:10.3389/fpsyg.2016.00279



Active Learning in Research Methods Classes Is Associated with Higher Knowledge and Confidence, Though not Evaluations or Satisfaction

Peter J. Allen* and Frank D. Baughman

School of Psychology and Speech Pathology, Curtin University, Perth, WA, Australia

OPEN ACCESS

Edited by:

Jason C. Immekus,
University of Louisville, USA

Reviewed by:

Xing Liu,
Eastern Connecticut State University,
USA

Dirk Van Rooy,
Australian National University,
Australia

*Correspondence:

Peter J. Allen
p.allen@curtin.edu.au

Specialty section:

This article was submitted to
Educational Psychology,
a section of the journal
Frontiers in Psychology

Received: 29 July 2015

Accepted: 12 February 2016

Published: 01 March 2016

Citation:

Allen PJ and Baughman FD (2016)
Active Learning in Research Methods
Classes Is Associated with Higher
Knowledge and Confidence, Though
not Evaluations or Satisfaction.
Front. Psychol. 7:279.
doi: 10.3389/fpsyg.2016.00279

Research methods and statistics are regarded as difficult subjects to teach, fueling investigations into techniques that increase student engagement. Students enjoy active learning opportunities like hands-on demonstrations, authentic research participation, and working with real data. However, enhanced enjoyment does not always correspond with enhanced learning and performance. In this study, we developed a workshop activity in which students participated in a computer-based experiment and used class-generated data to run a range of statistical procedures. To enable evaluation, we developed a parallel, didactic/canned workshop, which was identical to the activity-based version, except that students were told about the experiment and used a pre-existing/canned dataset to perform their analyses. Tutorial groups were randomized to one of the two workshop versions, and 39 students completed a post-workshop evaluation questionnaire. A series of generalized linear mixed models suggested that, compared to the students in the didactic/canned condition, students exposed to the activity-based workshop displayed significantly greater knowledge of the methodological and statistical issues addressed in class, and were more confident about their ability to use this knowledge in the future. However, overall evaluations and satisfaction between the two groups were not reliably different. Implications of these findings and suggestions for future research are discussed.

Keywords: active learning, research methods, statistics, computer based experiments, authentic data, canned data

INTRODUCTION

A cornerstone of educational practice is the notion that the more engaged the learner, the more interested, passionate and motivated they will become, and the better the outcome will typically be vis-à-vis their learning. This causal chain, of sorts, thus predicts that higher rates of student retention, better grades, and higher levels of satisfaction and enjoyment are more likely to follow when a student is genuinely curious and involved in their study. However, student engagement appears to be more difficult to achieve in some areas of study compared to others. For instance, within psychology, research methods and statistics are widely regarded as 'difficult' subjects to teach (e.g., Connors et al., 1998). Student attitudes toward these topics are often negative (Murtonen, 2005; Sizemore and Lewandowski, 2009), and their interest in them is low (Vittengl et al., 2004; Rottinghaus et al., 2006). This lack of engagement is likely to impact

student outcomes, contributing to poorer grades and higher rates of attrition. However, a basic understanding of research methods is essential in order for students to gain a fuller appreciation of the literature underpinning their later academic, or professional careers. Thus, there appears to be a clear and growing need to identify teaching strategies that are maximally effective at removing barriers to learning research methods. This view is echoed by recent calls to reform traditional methods for teaching research methods and statistics, and it finds support from recent research. For example, in the *Guidelines for Assessment and Instruction in Statistics Education* (GAISE; Aliaga et al., 2005) college report, published by the American Statistical Association, a number of recommendations are highlighted with regard to the teaching of statistics in higher education. These recommendations include emphasizing the development of statistical literacy and thinking, making use of real data, focusing on conceptual understanding (rather than procedures or formulae), promoting active learning, making use of technology and administering assessment appropriate to evaluating learning in the classroom.

The view that teaching research methods and statistics may require a particular kind of approach is further supported by a recent meta-analysis by Freeman et al. (2014). In their analysis, traditional methods of teaching statistics (e.g., lecturing to classes) was shown to be less effective in terms of student exam performance, and student satisfaction and enjoyment, compared to other subjects of study. The challenge facing teachers of statistics and research methods therefore is to make research methods more applied, relevant and engaging for students, whilst simultaneously improving students' understanding of statistics, their grades, and attendance rates (Hogg, 1991; Lovett and Greenhouse, 2000). In this article, we focus on the possible benefits of implementing two of the recommendations highlighted in the GAISE report. These are: (1) the use of real data, and (2) the use of an active learning methodology. We describe a study that examines the ways in which incorporating these recommendations into the teaching of research methods and statistics may positively affect student outcomes.

When applied to the teaching of research methods, active learning approaches typically involve students carrying out research, rather than merely reading about, or listening to instructors talk about it. Active learning in research methods and statistics classes may include taking part in demonstrations designed to illustrate methodological and statistical concepts, participating in authentic research, and working with data the students have been responsible for collecting. A great deal of work has explored the impact of active learning using 'hands-on' demonstrations of both statistical processes (e.g., Riniolo and Schmidt, 1999; Scitutto, 2000; Christopher and Marek, 2002; Fisher and Richards, 2004) and methodological concepts (e.g., Renner, 2004; Eschman et al., 2005; Madson, 2005). Importantly, the use of active learning methods in research methods and statistics appears to be successful at increasing levels of satisfaction and enjoyment and reducing failure rates (Freeman et al., 2014). Against this backdrop of findings, it might then seem reasonable to assume that the effects of active

learning would further contribute toward positive outcomes, for example on exam performance. However, this is not found to be the case. While students may report higher levels of enjoyment and usefulness of active learning demonstrations, these are not consistently associated with more beneficial learning outcomes (Elliott et al., 2010, though see also Owen and Siakaluk, 2011). Put another way, the subjective evaluation of one's enjoyment of a subject does not bear a direct relationship on the amount of knowledge acquired, or the extent to which one can apply knowledge in a given area (see e.g., Christopher and Marek, 2002; Copeland et al., 2010).

With regard to the use of real datasets in class exercises and assessments, this too has been proposed to hold a number of advantages (Aliaga et al., 2005). The advantages include: increased student interest; the opportunity for students to learn about the relationships between research design, variables, hypotheses, and data collection; the ability for students to use substantive features of the data set (e.g., the combination of variables measured, or the research question being addressed) as a mnemonic device to aid later recall of particular statistical techniques; and the added benefit that using real data can provide opportunities for learning about interesting psychological phenomena, as well as how statistics should be calculated and interpreted (Singer and Willett, 1990). Additionally, a number of studies have showed that when real, class-generated data are used students report higher levels of enjoyment, an enhanced understanding of key concepts, and are likely to endorse the use of real data in future classes (see e.g., Lutsky, 1986; Stedman, 1993; Thompson, 1994; Chapdelaine and Chapman, 1999; Lipsitz, 2000; Ragozzine, 2002; Hamilton and Geraci, 2004; Marek et al., 2004; Morgan, 2009; Neumann et al., 2010, 2013).

Overall, the benefits of using active learning and real data within research methods and statistics classes show much promise. However, to better understand how the implementation of these strategies results in positive outcomes, further empirical investigation is needed. First, we note a lack of research that has simultaneously targeted outcomes of satisfaction, evaluation and knowledge (i.e., performance). Each of these outcomes likely plays an important role in influencing student engagement. In this study we assess students on each of these components. Secondly, we eliminate a potential design confound that may have affected previous research, by ensuring highly similar contexts in both our intervention and our control group. The same instructors were used in both instances. In this way, we may be more confident that any effects we observe are more likely due to our manipulation (i.e., active learning versus control), than to student-instructor interactions.

Motivated by a desire to increase student engagement in our undergraduate statistics and research methods courses, we developed a series of activities for a 1.5-h workshop. In each of these activities, students participated in a computer-based psychological experiment, engaged in class discussions and activities around the methods used in the experiment, and then used data generated by the class to run a range of data handling and statistical procedures. In this paper, we describe an evaluation of the first of these workshop activities in terms of (a) its

subjective appeal to students; and, (b) its pedagogic effectiveness. It was hypothesized that, compared to control participants who were provided with the same content, but delivered using a didactic presentation and canned dataset, students who participated in the activity-based (active learning + real data) workshop would (H1) evaluate the workshop more favorably; (H2) report higher levels of satisfaction with the workshop; (H3) achieve higher scores on a short multiple-choice quiz assessing their knowledge of key learning concepts addressed in the workshop; and (H4) report significantly higher confidence about their ability to demonstrate skills and knowledge acquired and practiced in the workshop.

MATERIALS AND METHODS

Design

A non-equivalent groups (quasi-experimental) design was employed in this study, with intact tutorial classes randomly assigned to the two workshop versions. These workshop versions were equivalent in content, but differed in delivery format. The activity-based version of the workshop began with a computer-based experiment in which the students participated, and contained activities that required students to analyze data collected in class. The canned dataset version of the workshop differed in that it began with a short description of the computer-based experiment (presented by the same instructors as the activity-based workshop), but was otherwise equivalent to the activity-based workshop. As much as possible, the workshops were identical in all other respects. The independent variable in this study was workshop type, of which there were two levels: activity-based and didactic/canned. The four dependent variables were: (1) evaluations, (2) overall satisfaction, (3) knowledge, and (4) confidence.

Participants

Participants were recruited from a participant pool, within which students are required to participate in at least 10 points worth of research during each semester (or complete an alternate written activity). One point was awarded for participating in the current study. A total of 39 participants were obtained for final analysis. Initial comparisons between the activity-based group ($n = 25$; M age = 22.43, $SD = 4.95$; 68% female; M final grade = 61.12, $SD = 14.54$) and the didactic/canned group ($n = 14$; M age = 25.93, SD age = 12.27; 78.6% female; M final grade = 61.42, $SD = 11.90$) indicated that there were no reliable group differences in age, $t(15.59) = -1.22$, $p = 0.230$, $d = 0.37$, gender distribution, $\chi^2(1, N = 39) = 0.50$, $p = 0.482$, $\phi = 0.11$, or final semester grades, $t(36) = -0.066$, $p = 0.948$, $d = 0.02$.

This research complies with the guidelines for the conduct of research involving human participants, as published by the Australian National Health and Medical Research Council (The National Health, Medical Research Council, the Australian Research Council, and the Australian Vice-Chancellors' Committee [NH&MRC], 2007). Prior to recruitment of any participants, the study was reviewed and approved by the Human Research Ethics Committee at Curtin University.

Consent was indicated by the submission of an online evaluation questionnaire, as described in the participant information immediately preceding it.

Materials and Measures

Workshop

The activity-based version of the workshop commenced with students participating in a short computer based experiment designed to examine the effects of processing depth on recall. Class members were randomized to one of two processing conditions, *imagine* and *rehearse*, then asked to remember a list of 12 words presented on screen at a rate of one word every 2 s. Members of the *imagine* condition were encouraged to engage in deep processing by being instructed to "try to imagine each concept as vividly as possible such that you are able to remember it later." Members of the *rehearse* condition were encouraged to engage in shallow processing by being instructed to "try to rehearse each word silently such that you are able to remember it later." All students then completed multiplication problems for 150 s as a distractor task. Finally, all students were presented with 24 words, 12 of which were 'old' (i.e., appeared on the original list) and 12 of which were 'new'. They were asked to indicate whether each of the 24 words was 'old' or 'new' by pressing a relevant keyboard button.

This task was developed in Java by the second author, as existing commercial software packages were unsuitable for our purposes due to high annual licensing fees (e.g., St James et al., 2005), or an insufficient feature set (e.g., Francis et al., 2008). It was hosted on a private webserver, and accessed by students using a standard web browser (e.g., Firefox). The data generated by each student were saved to a MySQL database accessible to the class tutor from his/her networked workstation. Following their participation, students were provided with a brief written summary of the experiment, and asked to work together to address a series of questions about its key methodological features. These questions prompted students to identify and operationalize independent and dependent variables, write research and null hypotheses, visualize experimental designs using standard notation, and consider the purpose of randomization.

While the students worked on these questions, the tutor downloaded the class data and collated them into an SPSS data file that was subsequently uploaded to a network drive for students to access. After a brief class discussion around the methodology of the experiment, students were directed to open the SPSS data file, and commence work on a series of questions requiring various data handling techniques and statistical analyses to address. Specifically, students were required to identify the appropriate statistical test to compare the two conditions on classification accuracy, and then run, interpret and report (in APA style) an independent samples *t*-test (including assumption testing, and an effect size). The workshop concluded with a class discussion around the statistical analyses, findings and interpretation.

The didactic/canned version of the workshop was identical to the activity-based version, except it began with a short description of the computer based experiment (presented by the class tutor

with the aid of PowerPoint slides), and required students to analyze a canned data set, rather than class generated data.

Evaluation Questionnaire

The online evaluation questionnaire contained five sections, measuring the four DVs and capturing key demographic data. It is reproduced in full in the Appendix (available as Supplementary Material Data Sheet 1).

Section 1 (evaluations)

Section 1 of the online questionnaire contained 13 items assessing students' evaluations of the workshop. Although there are numerous measures that have been developed to allow students to evaluate units and courses, a review of the literature indicated that there are currently no instruments suitable for evaluating specific activities embedded within a unit or course. Consequently, this measure was developed specifically for the purposes of the current research (although inspired by the single-item measures that are frequently used in evaluations of teaching activities reported elsewhere). Participants responded to each item on a 7-point scale ranging from 1 (*Strongly disagree*) to 7 (*Strongly agree*), and examples of items on this measure include "this workshop was useful" and "this workshop was an effective way of teaching research methods and statistics." Although a small sample size limited our ability to examine the factor structure of this measure (for example, Pett et al. (2003), suggest a minimum of 10–15 cases per item for exploratory factor analysis), Cronbach's alpha was 0.96, indicating that it was internally consistent. Responses to the 13 items were summed to provide an overall index of how favorably students rated the workshop.

Section 2 (satisfaction)

The second section of the online questionnaire was a single item measure of overall satisfaction with the workshop, which respondents answered on a scale ranging from 1 (*Very Dissatisfied*) to 10 (*Very Satisfied*). The correlation between this single item measure and the sum of responses to the 13-item evaluation scale was $r = 0.91$, suggesting that they measured overlapping constructs.

Section 3 (knowledge)

Five multiple-choice questions were used to assess knowledge of the key learning outcomes addressed in the workshop. Each question provided four response options, of which only one was correct, thus total scores on this measure ranged from 0 to 5.

Section 4 (confidence)

This section of the questionnaire asked respondents to indicate on a 4-point scale ranging from 1 (*Not at all confident*) to 4 (*Very confident*) their confidence regarding their ability to apply seven specific skills developed in the workshop, assuming access to their notes and textbook. For example, "run and interpret and independent samples *t*-test using SPSS." Again, the small sample size limited our ability to examine the factor structure of this measure, although Cronbach's alpha was 0.84, indicating that it was internally consistent. Responses to the items on this measure were summed to provide an overall index of student confidence.

Section 5 (demographics)

The final section of the evaluation questionnaire asked students to specify their age, gender, and the day/time of the workshop they attended. The day/time information was used to assign participants to the levels of the independent variable.

Procedure

Before the start of semester, tutorial classes were block-randomized to the two workshop versions. The workshop was then delivered as part of the normal tutorial schedule. Participants were provided with an information sheet outlining the nature of the current study, and it was stressed that their involvement was (a) entirely voluntary, and (b) anonymous to the unit's teaching staff. At the end of the workshop, students were reminded about the research, and asked to complete the online evaluation questionnaire, which was linked from the unit's Blackboard site, within 48 h of the class finishing. Prior to accessing the online questionnaire, participants were presented with an online version of the information sheet hosted on our school website, as recommended by Allen and Roberts (2010).

RESULTS

Each hypothesis was tested with a Generalized Linear Mixed Model (GLMM), implemented via SPSS GENLINUX (version 22), with an alpha level of 0.0125 (to protect against the inflated risk of making Type 1 errors when conducting multiple comparisons on a single data set), and robust parameter estimation. GLMM is preferable to a series of independent samples *t*-tests or ordinary least squares (OLS) regression analyses, as it can accommodate dependencies arising from nested data structures (in this instance, 39 students nested in seven classes, facilitated by three tutors), non-normal outcome variables, and small, unequal group sizes. In each GLMM, there were two random effects (class and tutor)¹ and one fixed effect (condition) specified. A normal probability distribution was assumed for each outcome variable, and each was linked to the fixed effect with an identity function.

The fixed effects from the four GLMMs are summarized in **Table 1**, where it can be seen that members of the activity-based condition scored significantly higher than members of the didactic/canned condition on the knowledge and confidence measures, but not the evaluation and satisfaction measures. When indexed using Hedges' *g*, the knowledge and confidence effects could be characterized as 'large' and 'small,' respectively.

¹Note that for five of the eight tests of random effects, the variances were negative, and consequently set at zero during analyses. For iterative procedures (e.g., maximum likelihood estimation), this can occur when the variance attributable to a random effect is relatively small, and the random effect is having a negligible impact on outcome of the analyses. The remaining three random effects were non-significant, with Wald's *Z* ranging from 0.298 to 0.955 ($p = 0.765$ to 0.340). Despite their non-significance in the current context, the random effects of class and tutor were retained in our analyses, based on the common recommendation that non-independence of observations attributable a study's design ought to be routinely accounted for, regardless of the estimated magnitude of its impact (Murray and Hannan, 1990; Bolker et al., 2009; Thiele and Markussen, 2012; Barr et al., 2013).

TABLE 1 | Summary of differences between the two conditions on the four outcome variables.

Outcome	Estimated condition <i>M</i> (<i>SD</i>)		95% CI	<i>t</i>	<i>p</i>	<i>g</i>
	Activity-based	Didactic/canned				
Evaluations	4.77 (1.22)	4.71 (0.31)	[-0.59,0.72]	0.21	0.836	0.06
Satisfaction	6.68 (2.86)	7.00 (0.95)	[-2.02,1.37]	-0.38	0.703	0.13
Knowledge	3.72 (0.32)	3.43 (0.11)	[0.22,0.36]	8.07	<0.001	1.07
Confidence	2.38 (0.20)	2.33 (0.12)	[0.04,0.08]	5.86	<0.001	0.28

95% CI = 95% confidence interval of the difference between two means. *g* = Hedges' *g* for the weighted standardized difference between two means. *N* = 39 for all outcomes except Satisfaction, where *N* = 38.

DISCUSSION

We have focused on the implementation of two recommended strategies for teaching research methods and statistics: using real data, and following an active learning approach. Our results showed no reliable differences between groups in their rated evaluation of (H1), or satisfaction with (H2) the workshops. Those participants in the activity-based workshop were statistically no different in their views to those in the didactic/canned workshop. Indeed, it is interesting to note that both groups rated the workshops to be below-average (i.e., below the neutral-point) on the evaluation and satisfaction measures, suggesting that their views regarding the workshops were somewhere between ambivalent and negative. Overall, these findings were not as we predicted. Rather, we expected students in the activity-based workshop to find more satisfaction with their workshop and evaluate their learning experience more favorably. In-line with our predictions, however, was the finding that on the outcome measure of knowledge/performance, the activity-based group did significantly outperform those in the didactic/canned workshop (H3). Thus, while the groups did not differ in their apparent engagement, they nevertheless achieved different levels of knowledge. Also noteworthy, was the finding that the activity-based group were reliably different to the didactic/canned group in their reported levels of confidence to later apply the skills developed in the workshop (H4).

Seemingly, the results of this study sit at odds with the 'causal chain' we described in the introduction. One possible explanation is that for student satisfaction to be positively affected, students need to see the results of their engaged learning first, and perhaps these positive attitudes require time to accumulate. In our study, participants did not have this opportunity. A more interesting possibility is that rather than greater engagement being instrumental in promoting greater levels of satisfaction and enjoyment, which in turn promotes learning, that instead, one's level of satisfaction is in fact rather separate to the process of learning. If so, this would indicate that a combination of teaching strategies is needed to produce positive outcomes and student engagement. Accordingly, our results would be consistent with previous research that suggests exposure to research methods and statistics in an engaging environment can improve students' knowledge without necessarily affecting their attitudes (e.g., Sizemore and Lewandowski, 2009). This latter interpretation offers up a variety of potentially interesting research avenues. Minimally, the results of this study suggest against the tailoring

of content in educational curricular, based on the reported levels of satisfaction of students.

Limitations

While the results of the current study raise intriguing questions about the relationship between academic outcomes and self-reported student satisfaction and evaluations, it is important to note a number of possible limitations to the approach we took. The first of these concerns the relatively small, unequal number of participants in the activity-based (*n* = 25) versus canned/didactic (*n* = 14) groups. Clearly, to be more confident in our results, this study requires replication with a larger, more evenly spread sample. A second sampling limitation concerns the randomization of intact groups to conditions. Ideally, we would have randomized individual participants to either the activity-based or didactic/canned workshop, allowing for a true experimental test of each hypothesis. However, this was not possible due to the fact that students self-select into classes based on personal preferences and commitments.

A further possible limitation concerns the analytical approach we chose. Had we opted for another approach, for example independent samples *t*-tests, no reliable differences would have emerged (*ps* 0.385–0.839) and the implications of our study would be quite different. However, due to the fact that participants were recruited across a number of tutorial groups (*n* = 7) supervised by a number of instructors (*n* = 3), we deemed the use of GLMM procedures to be most appropriate. This is because GLMM is aptly suited to dealing with hierarchical data, and clustering effects that may have been present within nested groups of tutorials and instructors. GLMM has the further advantage over the *t*-test in that it may be more robust to dealing with unequal sample sizes (Bolker et al., 2009). Although our analysis showed no such clustering effects, in light of the sampling limitations, GLMM remained most suited to the data.

CONCLUSION

This paper describes the implementation and quasi-experimental evaluation of a relatively short (1.5 h) class activity in which students participated in an authentic computer-based psychological experiment, engaged in class discussion around its methods, and then used class-generated data to run a

range of data handling procedures and statistical tests. Results indicated that students who participated in this activity scored significantly higher than participants in a parallel didactic/canned class on measures of knowledge and confidence, but not on overall evaluations or satisfaction. In contrast to the view that student satisfaction is paramount in achieving positive learning outcomes, the results of the current study suggest that, at least during some points in the learning process, one's level of satisfaction has little effect. This would indicate that a combination of teaching strategies is needed to produce both positive outcomes and student engagement. Future research that employs large-scale, fully randomized experimental designs may have the best chance of revealing these strategies (Wilson-Doenges and Gurung, 2013).

AUTHORS CONTRIBUTIONS

PA conceived and designed the study and analyzed the data. PA and FB co-authored this manuscript. FB programmed the experimental task used as one level of the IV, wrote the documentation and spreadsheets used by the tutors to aggregate

the data for class use, and contributed to the overall design of the study.

FUNDING

This research was supported with an eScholar grant awarded to the first author by the Centre for eLearning, Curtin University, Australia.

ACKNOWLEDGMENTS

The authors would also like to acknowledge Dr. Robert T Kane from the School of Psychology and Speech Pathology at Curtin University for his statistical advice.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fpsyg.2016.00279>

REFERENCES

- Aliaga, M., Cobb, G., Cuff, C., Garfield, J., Gould, R., Lock, R., et al. (2005). *Guidelines for Assessment and Instruction in Statistics Education: College Report*. Alexandria, VA: American Statistical Association.
- Allen, P. J., and Roberts, L. D. (2010). The ethics of outsourcing survey research. *Int. J. Technoethics* 1, 35–48. doi: 10.4018/jte.2010070104
- Barr, D. J., Levy, R., Scheepers, C., and Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: keep it maximal. *J. Mem. Lang.* 68, 255–278. doi: 10.1016/j.jml.2012.11.001
- Bolker, B. J., Brooks, M. E., Clark, C. J., Geange, S. W., Poulsen, J. R., Stevens, M. H. H., et al. (2009). Generalized linear mixed models: a practical guide for ecology and evolution. *Trends Ecol. Evol.* 24, 127–135. doi: 10.1016/j.tree.2008.10.008
- Chapdelaine, A., and Chapman, B. L. (1999). Using community-based research projects to teach research methods. *Teach. Psychol.* 26, 101–105. doi: 10.1207/s15328023top2602_4
- Christopher, A. N., and Marek, P. (2002). A sweet tasting demonstration of random occurrences. *Teach. Psychol.* 29, 122–125. doi: 10.1207/S15328023TOP2902_09
- Conners, F. A., Mccown, S. M., and Roskos-Ewoldsen, B. (1998). Unique challenges in teaching undergraduate statistics. *Teach. Psychol.* 25, 40–42. doi: 10.1207/s15328023top2501_12
- Copeland, D. E., Scott, J. R., and Houska, J. A. (2010). Computer-based demonstrations in cognitive psychology: benefits and costs. *Teach. Psychol.* 37, 141–145. doi: 10.1080/00986281003626680
- Elliott, L. J., Rice, S., Trafimow, D., Madson, L., and Hipshur, M. F. (2010). Research participation versus classroom lecture: a comparison of student learning. *Teach. Psychol.* 37, 129–131. doi: 10.1080/00986281003626862
- Eschman, A., St James, J., Schneider, W., and Zuccolotto, A. (2005). PsychMate: providing psychology majors the tools to do real experiments and learn empirical methods. *Behav. Res. Methods* 37, 301–311. doi: 10.3758/BF03192698
- Fisher, L. A., and Richards, D. St. P. (2004). Random walks as motivational material in introductory statistics and probability courses. *Am. Stat. Assoc.* 58, 310–316. doi: 10.1198/000313004X5482
- Francis, G., Neath, I., and VanHorn, D. (2008). *CogLab on a CD: Version 2.0*, 4th Edn. Belmont, CA: Wadsworth.
- Freeman, S., Eddy, S. L., McDonough, M., Smith, M. K., Okoroafor, N., Jordt, H., et al. (2014). Active learning increases student performance in science, engineering, and mathematics. *Proc. Natl. Acad. Sci. U.S.A.* 111, 8410–8415. doi: 10.1073/pnas.1319030111
- Hamilton, M., and Geraci, L. (2004). Converting an experimental laboratory course from paper and pencil to computer. *Teach. Psychol.* 31, 141–143. doi: 10.1207/s15328023top3102_7
- Hogg, R. V. (1991). Statistical education: improvements are badly needed. *Am. Stat.* 45, 342–343. doi: 10.2307/2684473
- Lipsitz, A. (2000). Research methods with a smile: a gender difference exercise that teaches methodology. *Teach. Psychol.* 27, 111–113. doi: 10.1207/S15328023TOP2702_07
- Lovett, M. C., and Greenhouse, J. B. (2000). Applying cognitive theory to statistics instruction. *Am. Stat.* 54, 196–206. doi: 10.2307/2685590
- Lutsky, N. (1986). Undergraduate research experience through the analysis of data sets in psychology courses. *Teach. Psychol.* 13, 119–122. doi: 10.1207/s15328023top1303_4
- Madson, L. (2005). Demonstrating the importance of question wording on surveys. *Teach. Psychol.* 32, 40–43. doi: 10.1207/s15328023top3201_9
- Marek, P., Christopher, A. N., and Walker, B. J. (2004). Learning by doing: research methods with a theme. *Teach. Psychol.* 31, 128–131. doi: 10.1207/s15328023top3102_6
- Morgan, D. L. (2009). Using single-case design and personalized behavior change projects to teach research methods. *Teach. Psychol.* 36, 267–269. doi: 10.1080/00986280903175715
- Murray, D. M., and Hannan, P. J. (1990). Planning for the appropriate analysis in school-based drug-use prevention studies. *J. Consult. Clin. Psychol.* 58, 458–468. doi: 10.1037/0022-006X.58.4.458
- Murtonen, M. (2005). University students' research orientations: do negative attitudes exist toward quantitative methods? *Scand. J. Educ. Res.* 49, 263–280. doi: 10.1080/00313830500109568
- Neumann, D. L., Hood, M., and Neumann, M. M. (2013). Using real-life data when teaching statistics: student perceptions of this strategy in an introductory statistics course. *Stat. Educ. Res. J.* 12, 59–70.
- Neumann, D. L., Neumann, M. M., and Hood, M. (2010). The development and evaluation of a survey that makes use of student data to teach statistics. *J. Stat. Educ.* 18, 1–19.

- Owen, W. J., and Siakaluk, P. D. (2011). A demonstration of the analysis of variance using physical movement and space. *Teach. Psychol.* 38, 151–154. doi: 10.1177/0098628311411779
- Pett, M. A., Lackey, N. R., and Sullivan, J. J. (2003). *Making Sense of Factor Analysis: The Use of Factor Analysis for Instrument Development in Health Care Research*. Thousand Oaks, CA: Sage.
- Ragozzine, F. (2002). SuperLab LT: evaluation and uses in teaching experimental psychology. *Teach. Psychol.* 29, 251–254. doi: 10.1207/S15328023TOP2903_13
- Renner, M. J. (2004). Learning the Rescorla-Wagner model of Pavlovian conditioning: an interactive simulation. *Teach. Psychol.* 31, 146–148. doi: 10.1207/s15328023top3102_9
- Riniolo, T. C., and Schmidt, L. A. (1999). Demonstrating the gambler's fallacy in an introductory statistics class. *Teach. Psychol.* 26, 198–200. doi: 10.1207/S15328023TOP260308
- Rottinghaus, P. J., Gaffey, A. R., Borgen, F. H., and Ralston, C. A. (2006). Diverse pathways of psychology majors: vocational interests, self-efficacy, and intentions. *Career Dev. Q.* 55, 85–93. doi: 10.1002/j.2161-0045.2006.tb00007.x
- Sciutto, M. J. (2000). Demonstration of factors affecting the F ratio. *Teach. Psychol.* 27, 52–53. doi: 10.1207/S15328023TOP2701_12
- Singer, J. D., and Willett, J. B. (1990). Improving the teaching of applied statistics: putting the data back into data analysis. *Am. Stat.* 44, 223–230. doi: 10.2307/2685342
- Sizemore, O. J., and Lewandowski, G. W. Jr. (2009). Learning might not equal liking: research methods course changes knowledge but not attitudes. *Teach. Psychol.* 36, 90–95. doi: 10.1080/00986280902739727
- St James, J. D., Schneider, W., and Eschman, A. (2005). *PsychMate: Experiments for Teaching Psychology (Version 2.0)*. Pittsburgh, PA: Psychology Software Tools.
- Stedman, M. E. (1993). Statistical pedagogy: employing student-generated data sets in introductory statistics. *Psychol. Rep.* 72, 1036–1038. doi: 10.2466/pr0.1993.72.3.1036
- The National Health, Medical Research Council, the Australian Research Council, and the Australian Vice-Chancellors' Committee [NH&MRC]. (2007). *National Statement on Ethical Conduct in Human Research 2007 (Updated May 2015)*. Canberra, ACT: Author.
- Thiele, J., and Markussen, B. (2012). Potential of GLMM in modelling invasive spread. *CAB Rev.* 7, 1–10. doi: 10.1079/PAVSNR20127016
- Thompson, W. B. (1994). Making data analysis realistic: incorporating research into statistics courses. *Teach. Psychol.* 21, 41–43. doi: 10.1207/s15328023top2101_9
- Vittengl, J. R., Bosley, C. Y., Brescia, S. A., Eckardt, E. A., Neidig, J. M., Shelver, K. S., et al. (2004). Why are some undergraduates more (and others less) interested in psychological research? *Teach. Psychol.* 31, 91–97. doi: 10.1207/s15328023top3102_3
- Wilson-Doenges, G., and Gurung, R. A. R. (2013). Benchmarks for scholarly investigations of teaching and learning. *Aust. J. Psychol.* 65, 63–70. doi: 10.1111/ajpy.12011

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2016 Allen and Baughman. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

**Paper 4: Detecting Duplication in Students' Research Data: A Method and
Illustration**

Allen, P. J., Lourenco, A., & Roberts, L. D. (2016). Detecting duplication in students' research data: A method and illustration. *Ethics and Behavior*, 26, 300-311. doi:10.1080/10508422.2015.1019070

This is an Accepted Manuscript of an article published by Taylor & Francis in *Ethics and Behavior* on 26 Feb 2015, available online: <http://www.tandfonline.com/10.1080/10508422.2015.1019070>

In compliance with Taylor & Francis's terms and conditions of use, this is the post-print version of the abovementioned work. Readers wishing to cite this paper are encouraged to source the final published version, available at the URL above.

Detecting Duplication in Students' Research Data: A Method and Illustration

Peter J. Allen, Amanda Lourenco, and Lynne D. Roberts

Curtin University, Australia

Author Note

Peter J. Allen, School of Psychology and Speech Pathology, Curtin University, GPO Box U1987, Perth, Western Australia, 6845. p.allen@curtin.edu.au **

Amanda Lourenco, School of Psychology and Speech Pathology, Curtin University, GPO Box U1987, Perth, Western Australia, 6845. amanda.lourenco@student.curtin.edu.au

Lynne D. Roberts, Associate Professor, School of Psychology and Speech Pathology, Curtin University, GPO Box U1987, Perth, Western Australia, 6845. lynne.roberts@curtin.edu.au

This research was supported by a grant from the Research and Development Committee of the School of Psychology and Speech Pathology, Faculty of Health Sciences, Curtin University.

** Corresponding author.

Abstract

Research integrity is core to the mission of higher education. In undergraduate student samples, self-reported rates of data fabrication have been troublingly high. Despite this, no research has investigated undergraduate data fabrication in a more systematic manner. We applied duplication screening techniques to 18 data sets submitted by psychology honours students for assessment. Although we did not identify any completely duplicated cases, there were numerous partial duplicates. Rather than indicating fabrication however, these partial duplicates are likely a consequence of poor measure selection, insufficient data screening and/or participant characteristics. Implications for the teaching and supervision of honours students are discussed.

Keywords: data fabrication; data duplication; research, ethics

Detecting Duplication in Students' Research Data: A Method and Illustration

Academic integrity is core to the mission of the higher education sector. Research misconduct, which has been widely defined as “fabrication, falsification, or plagiarism in proposing, performing, or reviewing research, or in reporting research results” (Office of Science and Technology Policy, 2000), fundamentally undermines this mission. Of these three forms of research misbehaviour, fabrication (making up data) and falsification (changing data) are often seen as most problematic, as they directly distort scientific knowledge and the decisions based on it (Steneck, 2006). Both have received considerable media attention in recent years (Anderson, Shaw, Steneck, Konkle, & Kamata, 2013).

For example, data fabrication in biomedical research hit the headlines in 2005/6 when it was revealed that South Korean human stem-cell researcher, Woo Suk Hwang, had invented much of the data on which two landmark papers published in *Science* (and later retracted; Hwang et al., 2004, 2005) were based (Cyranoski, 2006). Australian obstetrician William McBride was found guilty of data falsification in research he claimed demonstrated that Debendox, a morning sickness drug, caused birth deformities (Milliken, 1993). Despite evidence suggesting that cases like these occur most commonly in biomedical research (Fanelli, 2009; Grieneisen & Zhang, 2012; Stroebe, Postmes, & Spears, 2012), our own field, psychology, has not escaped scandal. In one recent example, Dutch social psychologist, Diederik Stapel, was exposed as having published over 50 papers based on fraudulent data (Stroebe et al., 2012).

Although it is tempting to conclude that cases like these represent isolated ‘bad apples’ in an otherwise honest system, there is evidence to indicate that the problem is somewhat more widespread. For example, Steneck (2006) calculated that around .001%

of scientists in the US are found guilty of misconduct by a federal oversight agency each year. Based on the ratio of retractions to articles indexed in the PubMed database, Claxton (2005) estimated that between .002% and .02% of published papers report fraudulent data. In a comprehensive review of retraction notices published in 42 scholarly databases between 1980 and 2010, Grieneisen and Zhang (2012) found that retraction rates ranged from .02% to 5.62% across 1,796 unique journal titles, and had increased rapidly in the last 10 years (see also, Steen, 2012). Questionable data or interpretations (including data fabrication and falsification) were cited as the reasons for retraction in 43% of the cases they observed (Grieneisen & Zhang, 2012). Similarly, Fang, Steen, and Casadevall (2012) reported 67.4% of 2047 retractions of articles indexed by PubMed were attributable to misconduct, with 43.4% identified or suspected as fraudulent. Although troubling, these figures almost certainly underestimate the true scope of the problem, considering that rates of confirmed misconduct are a poor proxy for actual rates of misconduct, and cases that are reported and investigated likely only represent 'the tip of the iceberg' (Steneck, 2006).

When self-report data are used to estimate research misconduct rates, the figures are somewhat higher. For example, in a survey funded by The Office of Research Integrity in the US, the Gallup Organization (2008) asked just one randomly selected principal investigator at each of over 4,000 unique schools/departments to report on the research misconduct they had observed in their own workplaces during the previous three years. With a response rate of over 50%, they estimated that around 1.5% of research conducted in the US each year involves some form of serious misconduct. They further estimated that around 60% of that misconduct involves either data fabrication or falsification (Gallup Organization, 2008). In a recent meta-analysis of 18

surveys, Fanelli (2009) found that 1.97% of scientists admitted to having fabricated or falsified data at least once, while nearly 15% reported observing colleagues engaged in this practice. Although these figures almost certainly include some duplicate cases, with the same instances of misconduct being reported by multiple researchers on some occasions (Strobe et al., 2012), Fanelli (2009) argued that they are probably still conservative estimates of the true prevalence of research misconduct within the scientific community. This conclusion is supported by at least three lines of evidence. First, people routinely underreport their own criminal and socially sensitive/undesirable behaviours in self-report surveys (e.g., Farrington, 2001; Krumpal, 2013; Magura & Kang, 1996; Tourangeau & Yan, 2007). This is especially so when social expectations are inconsistent with the behaviours under investigation. This is the case in science, where researchers are expected to act with integrity, a value wholly inconsistent with data fabrication and falsification. Second, researchers report a much higher degree of 'willingness' to engage in future misconduct than they report actual past misconduct (e.g., Eastwood, Derish, Leash, & Ordway, 1996). Third, when researchers are provided with incentives for honesty (e.g., a donation to a charity of their choice, calculated dependent on the estimated truthfulness of their answers), they self-report higher levels of questionable research practices (John, Loewenstein, & Prelec, 2012). John and colleagues (2012) argue that prevalence estimates derived from these methods are likely to be more valid than straight self-reports, a conclusion which is hinged on knowledge that people are less likely to report doing socially sensitive things they haven't done, than not report things they have done (Krumpal, 2013).

Although rates of data fabrication and falsification amongst professional scientists are concerning, they pale into insignificance when compared to self-reported

rates of serious research misconduct amongst university students. For example, some 19% of McCabe's (2005) online sample of over 46,000 North American undergraduate students self-reported fabricating or falsifying laboratory data at least once in the previous year. Eight percent admitted to falsifying research data during the same time period. Amongst his sample of over 7,000 graduate students, the rates of falsifying laboratory and research data were 7% and 4% respectively (McCabe, 2005). A large Australian study by Brimble and Stevenson-Clarke (2005) produced broadly comparable findings, with over 21% of nearly 1,200 (predominantly undergraduate) students indicating that they had falsified the results of their own research at least once. This rate was several times higher than the prevalence estimates provided by academic staff at the same universities. Furthermore, and in contrast to the academic staff, the majority of the students saw falsification as only 'minor cheating' at worst.

In several smaller studies, the numbers are even more troubling. For example, over 60% of Franklyn-Stokes and Newstead's (1995) sample of UK undergraduate science students confessed to inventing or altering research data in the past. Sixty-seven percent of Lawson, Lewis, and Birk's (1999/2000) US sample of biology, chemistry and anatomy undergraduates indicated that they manipulate or make up data at least 'sometimes'. Finally, nearly everyone in Davidson, Cate, Lewis, and Hunter's (2000) US sample of undergraduate biology and chemistry students admitted to manipulating data at least 'often', with many indicating that they do it 'almost always'. The most common reason cited for manipulating data was to obtain a better grade (Davidson et al., 2000).

Although these findings are quite diverse, and varied operational definitions and methodologies make it difficult to pin-down the exact extent of data fabrication and

falsification amongst undergraduate students, when looked at in combination, they nevertheless suggest that these are reasonably common practices throughout large segments of the student population. This is concerning for a number of reasons. For example, cheating “threatens the equity and efficacy of instructional measurement” (Brimble & Stevenson-Clarke, 2005, p. 20), as it can artificially inflate the grades of students who cheat, relative to those who do not. This is especially pertinent in the natural sciences, where producing data that ‘support the hypothesis’ is often directly rewarded with higher grades (Lawson et al., 1999/2000). Furthermore, cheating may result in impoverished learning, leaving students less equipped to deal with more advanced topics in their chosen subject areas (Brimble & Stevenson-Clarke, 2005). Beyond such consequences at the individual and institutional levels, cheating also has broader societal implications. For example, broad awareness of cheating may undermine public trust in entire professions (Marsden, Carroll, & Neill, 2005). For instance, it is not difficult to imagine some people becoming wary of visiting all doctors in the weeks following publication of news stories exposing cheating at a local medical school. Furthermore, dishonesty in college is known to correlate strongly with dishonesty in the workplace (e.g., partial $r > .60$ in Nonis & Swift, 2001), which suggests that dishonest behaviours learned or practiced while studying may transfer or generalise to other contexts. Finally, these findings are all derived from self-report data, which are subject to selection, social desirability and other biases (Crown & Spiller, 1998; Macfarlane, Zhang, & Pun, 2012), and thus may actually underestimate the true scope of academic misconduct in the university student population.

Consequently, a few researchers have attempted to measure student academic misconduct via more objective methods. For example, Pullen, Ortloff, Casey, and Payne

(2000) collected 62 'cheat sheets' they found discarded around their university campus, which they then content analysed. They found that business students tended to be over-represented in their sample, and the 'typical' cheat sheet was small enough to easily conceal in the palm of a hand, and densely packed with organised lists of facts. Ward and Beck (1990) asked students to self-score multiple-choice exams in class, and then compared these self-scores to the students' actual performance on the exam (as determined by Scantron scoring). Some 28% of the sample self-scored themselves higher than their actual scores. By way of contrast, no students gave themselves a score lower than their actual score. Karlins, Michaels, and Podlogar (1988) compared students' written assignments to those submitted in a previous semester, and found that around three percent had been 'recycled' (either by the same, or a different student). Finally, Martin, Rao, and Sloan (2009) reviewed business administration students' Turnitin similarity reports to detect plagiarism on written assignments, a practice that is now common with Turnitin reporting use of their text matching software in more than 3,500 higher education institutions (iParadigms, 2013). After Martin and colleagues (2009) screened each report to ensure the absence of any false-positives, they found that 61% of their sample met their threshold for plagiarism (at least three percent of the assignment matching a source in the Turnitin database). While these studies provide interesting insights into how students may cheat in some narrowly defined circumstances, they shed little light on the broader scope of the problem. Furthermore, convenience sampling and the esoteric contexts in which they were conducted make generalisation difficult. Finally, we are not aware of any research that has attempted to objectively measure practices suggestive of data fabrication and falsification within a student sample.

There are, however, a number of statistical methods that are available for revealing the possibility of such practices. For example, Evans (2001) describes various univariate, bivariate and multivariate techniques (statistical and graphical) that are useful for screening data derived from clinical trials. Similar methods are discussed by Buyse and colleagues (1999), and illustrated with both fraudulent and legitimate data in Al-Marzouki, Evans, Marshall, and Roberts (2005). Whilst valuable, many of these techniques rely on comparing the characteristics of randomised groups at baseline, which limits their applicability to experimental research for which pre-intervention data are available. When it comes to other types of research data, the literature is less well developed.

Nevertheless, some of the techniques described in Evans (2001) and elsewhere do have broader applicability. One such technique has been developed by Blasius and Thiessen (2012), and can be used to quickly screen data sets for either partially or completely duplicated cases, which may indicate data fabrication. This technique relies on Principal Components Analysis (PCA), and after establishing the rationale for the current research, we will describe its application to the sort of data sets typically produced by researchers (including student researchers) working in the behavioural and social sciences.

The Current Research

In psychology, the honours program is the primary route to both professional (e.g., clinical, organisational, counselling psychology etc.) and research careers, and often provides the first opportunity for students to conceptualise a research project and collect a substantial amount of raw research data. Practices developed during honours may continue into postgraduate studies, and working life (Nonis & Swift, 2001). We are

not aware of any published research on either the self-reported rates of data fabrication or falsification amongst honours students, or the systematic analysis of honours students' raw data for the presence of possible indicators of fabricated or falsified data. The recent development of statistical techniques for detecting certain patterns of data within a broad range of datasets (Blasius & Thiessen, 2012) makes such analysis possible.

The aim of the current research was to systematically examine a sample of psychology honours students' data sets for characteristics which could be attributable to data fabrication. The specific characteristic in question was the presence of either partially or completely duplicated cases. The two research questions driving this research were: (1) is there any evidence of data duplication in the data sets students present with their honours dissertations? (2) If any duplication is found, what are its likely causes?

Method

Participants

In 2012, 32 psychology honours students submitted dissertations in our school, which is situated in a medium-sized, medium-ranked Australian university. Of these, 12 students worked with existing data sets, or collected qualitative data, and were thus excluded from the current sample. A further two students were excluded, as one did not include raw data as a dissertation appendix, and the other's data set was very small, which would have prevented running the analyses described below. Therefore, we analysed a final convenience sample of 18 psychology honours students' dissertations and raw quantitative data sets.

In our school, completed honours dissertations (including raw data sets, which are typically included as digital appendices to each dissertation) are collected in our library, and thus become a part of the public record. Consequently, informed consent was not required from students prior to including their work in our sample (and seeking it would have undermined the intention of the research). Before examining any data sets, the second author (who was not familiar with the 2012 honours students) anonymised them, by giving each a randomly generated file name, and re-naming each variable in each file with a generic code (excluding age and gender). Before examining any dissertations, they were reduced and anonymised by a research assistant. This involved extracting only those sections relevant to data analyses and measurement, and then hashing out any variable names. Together, these processes ensured that we would be unable to link dissertations or data sets back to specific students in the event that any possible fabrication was suspected, and that there could be no adverse consequences to students as a result of their work being included in this research. This research conformed to the guidelines for ethical conduct in human research articulated by the Australian National Health and Medical Research Council (2007), and was approved by our local Human Research Ethics Committee prior to commencing.

Materials and Procedure

A typical honours level data file will contain at least 40 to 50 variables (and often many more), with a variety of response formats (e.g., dichotomous, 5-point Likert, 7-point Likert etc.). The probability of two identical strings of data more than several items long occurring due purely to chance within one of these files is extremely small. For example, if we were to take a 10-item segment of a given data file, and each of those 10-items used a 6-point response format, there are theoretically 10^6 (i.e., 1

million) possible permutations of responses. Consequently, the likelihood of two or more conscientious respondents producing exactly the same permutation of 10 responses within a small honours data file (typically no longer than a couple of hundred cases) due to chance is extremely small. Therefore, it is reasonable to conclude that the presence of such duplicates is probably not due to chance. Furthermore, if there are several sets of duplicates, further investigation to ascertain their likely causes is certainly warranted. These are the basic assumptions on which the following procedure (as first described by Blasius & Thiessen, 2012) was based.

First, we took variables 6 through 15 of each data file (assuming that if deliberate duplication had occurred and the student had attempted to hide their actions, the first and/or last several items of the data file are those most likely to have been modified) and calculated the maximum number of theoretically possible permutations of responses. If the maximum number of permutations possible was below 1 million, successive variables were sampled until this threshold was reached. Next, we subjected the sampled variables to PCA in SPSS (version 20), and saved the component scores for the first component extracted. Note that strings of identical (duplicate) responses will yield identical component scores. (Indeed, this will be the case, regardless of which extraction method is used.) Third, the component scores were graphed, as illustrated in Figure 1. This allowed us to quickly identify whether or not any cases shared component scores. If each case had a unique component score (as illustrated in the left panel of Figure 1), there was no evidence of duplication, and the analysis stopped. However, if any component score was shared by two or more cases (as can be seen in the right panel of Figure 1), they were flagged as ‘potentially problematic cases’ (PPCs), and a similarity index was recorded for each. This similarity index was

calculated by dividing the number of variables the cases were identical on by the total number of variables in the data file, and then multiplying by 100 to derive a percentage. In situations where three or more cases had the same component score, pairs of cases were compared, and only the highest similarity index was recorded against each. The process described above was then repeated from the other end of the data file, to ensure that all PPCs were captured.

<insert Figure 1 about here>

To investigate the possible causes of the identified PPCs in greater detail, the data sets were systematically examined with reference to the dissertation extracts. Specifically, we studied the patterns of responses across each measure in each data set, and documented the nature of each instance where PPCs clustered. Each instance was then cross-referenced against the relevant dissertation to determine whether or not the student author had identified the issue and, if so, the reasons they attributed it to.

Results

The percentage of PPCs in each of the 18 analysed student data files ranged from 0% through to 46.92%, as illustrated by the bars in Figure 2. Sixteen of the 18 data files contained at least some PPCs, but none contained more than 50%. Furthermore, none of the data sets contained fully duplicated cases (i.e., cases with a 100% similarity index), although three contained cases with similarity indices in excess of 90% (as illustrated by the solid line in Figure 2), and 15 contained cases with similarity indices above 50% (as illustrated by the dotted line).

<insert Figure 2 about here>

When the data sets were individually examined with reference to their corresponding dissertations, most PPCs were characterised by one of two features: (1)

long runs of responses to items within a multiple-item measure at either the ceiling or floor; or (2) patterned responses to items on one or more multiple-item measures. As evidence for (1), the majority of items on six different measures had medians at either the lowest or highest response option (range = 50% to 77% of items). A seventh measure had extreme medians on 22% of items. All seven of these measures contained at least one item with more than 70% of all responses at either the ceiling or floor. Regarding (2), eight cases of patterned responding (e.g., 2-2-2-2 or 2-4-2-4) were identified. In the worst instance, a full 10.6% of respondents to one 8-item measure provided an identical response to every item. Only two students identified any problems associated with floor/ceiling effects or patterned responding in their dissertations. The first indicated that one case was removed prior to analysis due to patterned responding, while the second reported on the extreme skew of some item level data.

Discussion

Using the methods described by Blasius and Thiessen (2012), we examined 18 data files submitted with dissertations for assessment by psychology honours students for the existence of partially or completely duplicated cases. Although we did not identify any completely duplicated cases, there were many partial duplicates. Partial duplicates are pairs or sets of cases with matching strings of data that are too long to be reasonably attributable to chance. There are a number of possible explanations for these findings, the majority of which do not suggest any nefarious intent on the part of the students responsible for collecting the data.

First, it may suggest poor measure selection by student researchers. Closer examination of each data file revealed several which were plagued by stereotypical response sets (e.g., responding with a series of 1s on a multiple-item Likert scale). Such

response sets could result from using a measure that is unsuitable for the target population (e.g., a measure designed to assess the severity of schizophrenic symptoms being used with a non-clinical population), or a measure containing items that were difficult to comprehend or irrelevant for sections of the target population.

Second, it may suggest participant fatigue, carelessness and/or socially desirable responding, particularly in online survey-based studies administered through a human subject pool, where participants were identifiable (to permit allocation of course credit) and sometimes asked in excess of 130 questions. Research indicates that this combination of factors may be particularly conducive to careless responding (Meade & Craig, 2012), and that identified surveys tend to elicit higher levels of socially desirable responding than anonymous surveys (Dodou & de Winter, 2014). No students identified these issues as potential concerns within their dissertations.

Third, it may suggest insufficient data screening by student researchers, who ought to have removed cases that had obviously not responded conscientiously (e.g., participants with very short completion times who selected the middle response option for virtually every question) prior to running hypothesis tests. Failure to do so introduces unnecessary error variance into data, which reduces statistical power and the likelihood of detecting meaningfully sized effects (Maniaci & Rogge, 2014). It should be noted that such carelessness and/or incompetence, whilst not representing research misconduct, may still be considered unethical (Wasserman, 2013).

Finally, the partially duplicated cases that we observed could also suggest fabrication coupled with a small amount of data point adjustment, whereby one or two data points were modified post-duplication to ensure that the fabricated cases were not completely identical. All three data sets with similarity indices above 90% contained

cases that differed only on age and/or gender, and one other variable unique to the data set in question.

It should be noted that the explanations presented above are somewhat speculative. It is impossible to fully establish their veracity without interviewing or surveying the researchers and participants involved in each study. It is also important to note that we only considered one specific type of data fabrication (duplication by means of cutting-and-pasting), and so cannot discount the possibility that the student researchers in our sample engaged in other, subtler, fraudulent behaviours. Methods to detect these fraudulent behaviours are documented elsewhere (e.g., Buyse et al., 1999; Evans, 2001). Furthermore, the absence of any egregious data duplication in this small ($N = 18$) convenience sample of Australian psychology honours students should not be used to conclude that undergraduate researchers elsewhere do not engage in such practices. Indeed, several self-report studies suggest that such behaviour may be more common in the natural and physical sciences, where there is a greater expectation that hypotheses will be supported, particularly in straight replications of prior studies (Davidson et al., 2000; Franklyn-Stokes & Newstead, 1995; Lawson et al., 1999/2000). This is an obvious avenue for future research, which should aim to investigate the prevalence of data duplication in larger samples from a variety of disciplines and institutions. Such research should also consider additional plausible causes and correlates of data duplication, including supervisor experience, student ability and various contextual factors. Indeed, one such factor that may have reduced the tendency of this particular cohort of psychology honours students engage in questionable research practices, relative to those from previous of successive years, was the salience of the Diederik Stapel case at the time (see Stroebe et al., 2012). In fact, an article authored by

Stapel was the focus of a major assignment in an advanced research methods course, which many members of the sample completed concurrently to their dissertation research. The article was retracted by the publisher shortly before the assignment submission date, resulting in extended class discussion on academic integrity and research misconduct.

Furthermore, future research should consider methods of assessing the reliability, sensitivity and validity of the techniques described in this paper, which we have not sufficiently addressed. For example, if the method is valid it should be able to detect duplicate cases randomly inserted into a random subset of existing data files by a second, independent researcher. Alternatively, if duplicate cases are an indicator of misconduct, and if students more likely to engage in misconduct are also more likely to engage in other ethically questionable academic behaviours (e.g., Broeckelman-Post, 2009), then dissertations have been identified as problematic for other reasons (e.g., plagiarism) should have a higher probability of also containing duplicate cases than those which have been passed without concern.

Despite the above limitations, there are a number of pedagogic implications that emerge from this research, relating particularly to the teaching and supervision of honours students across disciplines. First, if poor measure selection is responsible for a lack of variability in student data, efforts can be made in class to stress the importance of selecting measures that are valid for the population with which they will be used, and pre-testing them thoroughly in times of doubt. Similar efforts can be made to emphasise the importance of brevity when selecting measures for survey research, particularly if the intention is to administer them online (either anonymously or through subject pool management software), in order to reduce the likelihood of respondent fatigue and

careless responding. Furthermore, time can be spent with students discussing the conceptual importance of data screening and cleaning, as well as the actual mechanics of performing these tasks. Additionally, a number of studies also recommend explicitly addressing research ethics through policies and mentoring (Fisher, Fried, & Feldman, 2009; Fisher, Fried, Goodman, & Germano, 2009). Reminding students about research misconduct has been found to significantly reduce cheating (e.g., McCabe, Trevino, & Butterfield, 2001). Finally, talking with students about the importance of non-significant findings (Ionnadis, 2005; Schooler, 2011) may also help to relieve some of the pressure that many feel to 'reject the null hypothesis'.

In conclusion, this paper describes and illustrates a technique that can be quickly applied to a wide variety of data sets to detect the presence of partially or completely duplicated cases. When it was applied to 18 data sets submitted by psychology honours students for assessment, no complete duplicates were identified, although there were numerous partial duplicates. These duplicates may indicate data fabrication, although in the current context, a more benign etiology appears likely.

References

- Al-Marzouki, S., Evans, S., Marshall, T., & Roberts, I. (2005). Are these data real? Statistical methods for the detection of data fabrication in clinical trials. *BMJ*, *331*, 267-270. doi:10.1136/bmj.331.7511.267
- Anderson, M. S., Shaw, M. A., Steneck, N. H., Konkle, E., & Kamata, T. (2013). Research integrity and misconduct in the academic profession. In M. B. Paulsen (Ed.), *Higher education: Handbook of theory and research* (Vol. 28, pp. 217-261). Amsterdam, Netherlands: Springer. doi:10.1007/978-94-007-5836-0_5
- Blasius, J., & Thiessen, V. (2012). *Assessing the quality of research data*. London, UK: Sage.
- Brimble, M., & Stevenson-Clarke, P. (2005). Perceptions of the prevalence and seriousness of academic dishonesty in Australian universities. *The Australian Educational Researcher*, *32*(3), 19-44. doi:10.1007/BF03216825
- Broeckelman-Post, M. A. (2009). *Building a culture of academic integrity: The role of communication in creating and changing understandings and enactments of academic integrity* (Doctoral dissertation). Retrieved from http://rave.ohiolink.edu/etdc/view?acc_num=ohiou1242313551
- Buyse, M., George, S. L., Evans, S., Geller, N. L., Ranstam, J., Scherrer, B., ... Verma, B. L. (1999). The role of biostatistics in the prevention, detection and treatment of fraud in clinical trials. *Statistics in Medicine*, *18*, 3435-3451. doi:10.1002/(SICI)1097-0258(19991230)18:24<3435::AID-SIM365>3.0.CO;2-O

- Crown, D. F., & Spiller, M. S. (1998). Learning from the literature on collegiate cheating: A review of empirical research. *Journal of Business Ethics, 17*, 683-700. doi:10.1023/A:1017903001888
- Cyranoski, D. (2006). Verdict: Hwang's human stem cells were all fakes. *Nature, 439*, 122-123. doi:10.1038/439122a
- Davidson, E. W., Cate, H. E., Lewis, C. M., Jr., & Hunter, M. (2000, November 19-21). *Data manipulation in the undergraduate laboratory: What are we teaching?* Paper presented at the First ORI Research Conference of Research Integrity, Bethesda, MD. Retrieved from <http://ori.dhhs.gov/documents/papers/davidson.pdf>.
- Dodou, D., & de Winter, J. C. F. (2014). Social desirability is the same in offline, online and paper surveys: A meta-analysis. *Computers in Human Behavior, 36*, 487-495. doi:10.1016/j.chb.2014.04.005
- Eastwood, S., Derish, P., Leash, E., & Ordway, S. (1996). Ethical issues in biomedical research: Perceptions and practices of postdoctoral research fellows responding to a survey. *Science and Engineering Ethics, 2*, 89-114. doi:10.1007/BF02639320
- Evans, S. (2001). Statistical aspects of the detection of fraud. In S. Lock, F. Wells, & M. J. G. Farthing (Eds.), *Fraud and misconduct in biomedical research* (3rd ed., pp. 186-204). London, UK: BMJ Books.
- Fanelli, D. (2009). How many scientists fabricate and falsify research? A systematic review and meta-analysis of survey data. *PLOS One, 4*(5), e5738. doi:10.1371/journal.pone.0005738

- Fang, F. C., Steen, R. G., & Casadevall, A. (2012). Misconduct accounts for the majority of retracted scientific publications. *Proceedings of the National Academy of Sciences of the United States of America*, *109*, 17028-17033. doi:10.1073/pnas.1212247109
- Farrington, D. P. (2001). *What has been learned from self-reports about criminal careers and the causes of offending?* Home Office Online Report retrieved from http://www.crim.cam.ac.uk/people/academic_research/david_farrington/srdrep.pdf
- Fisher, C. B., Fried, A. L., & Feldman, L. G. (2009). Graduate socialization in the responsible conduct of research: A national survey on the research ethics training experiences of psychology doctoral students. *Ethics & Behavior*, *19*, 496-518. doi:10.1080/10508420903275283
- Fisher, C. B., Fried, A. L., Goodman, S. J., & Germano, K. K. (2009). Measures of mentoring, department climate, and graduate student preparedness in the responsible conduct of psychological research. *Ethics & Behavior*, *19*, 227-252. doi:10.1080/10508420902886726
- Franklyn-Stokes, A., & Newstead, S. E. (1995). Undergraduate cheating: who does what and why? *Studies in Higher Education*, *20*, 159-172. doi:10.1080/03075079512331381673
- Grieneisen, M. L., & Zhang, M. (2012). A comprehensive survey of retracted articles from the scholarly literature. *PLOS One*, *7*(10), e44118. doi:10.1371/journal.pone.0044118

- Gallup Organization. (2008). *Final report: Observing and reporting suspected misconduct in biomedical research*. Retrieved from http://ori.hhs.gov/sites/default/files/gallup_finalreport.pdf
- Hwang, W. S., Roh, S. I., Lee, B. C., Kang, S. K., Kwon, D. K., Kim, S., ... Schatten, G. (2005). Patient-specific embryonic stem cells derived from human SCNT blastocysts. *Science*, *308*, 1777-1783. doi:10.1126/science.1112286
- Hwang, W. S., Ryu, Y. J., Park, J. H., Park, E. S., Lee, E. G., Koo, J. M., ... Moon, S. Y. (2004). Evidence of a pluripotent human embryonic stem cell line derived from a cloned blastocyst. *Science*, *303*, 1669-1674. doi:10.1126/science.1094515
- Ionnadis, J. P. A. (2005). Why most published research findings are false. *PLoS Medicine*, *2*(8), e124. doi: 10.1371/journal.pmed.0020124
- iParadigms, LLC. (2013). *Turnitin: Colleges & university*. Retrieved 10 August, 2013, from http://turnitin.com/en_us/customers/college-and-university
- Karlins, M., Michaels, C., & Podlogar, S. (1988). An empirical investigation of actual cheating in a large sample of undergraduates. *Research in Higher Education*, *29*, 359-364. doi:10.1007/BF00992776
- Krumpal, I. (2013). Determinants of social desirability bias in sensitive surveys: A literature review. *Quality and Quantity*, *47*, 2025-2047. doi:10.1007/s11135-011-9640-9
- Lawson, A. E., Lewis, C. M., Jr., & Birk, J. P. (1999/2000). Why do students “cook” data? A case study on the tenacity of misconceptions. *Journal of College Science Teaching*, *29*, 191-198.

- Magura, S., & Kang, S. -Y. (1996). Validity of self-reported drug use in high risk populations: A meta-analytical review. *Substance Use & Misuse, 31*, 1131-1153. doi:10.3109/10826089609063969
- Maniaci, M. R., & Rogge, R. D. (2014). Caring about carelessness: Participant inattention and its effects on research. *Journal of Research in Personality, 48*, 61-83. doi:10.1016/j.jrp.2013.09.008
- Martin, D. E., Rao, A., & Sloan, L. R. (2009). Plagiarism, integrity, and workplace deviance: A criterion study. *Ethics & Behavior, 19*, 36-50. doi:10.1080/10508420802623666
- Marsden, H., Carroll, M., Neill, J. T. (2005). Who cheats at university? A self-report study of dishonest academic behaviours in a sample of Australian university students. *Australian Journal of Psychology, 57*, 1-10. doi:10.1080/00049530412331283426
- Macfarlane, B., Zhang, J., & Pun, A. (2012). Academic integrity: A review of the literature. *Studies in Higher Education*. Advance online publication. doi:10.1080/03075079.2012.709495
- McCabe, D. L. (2005). Cheating among college and university students: A North American perspective. *International Journal of Educational Integrity, 1*(1). Retrieved from <http://www.ojs.unisa.edu.au/index.php/IJEI/article/view/14>
- McCabe, D. L., Trevino, L. K., & Butterfield, K. D. (2001). Cheating in academic institutions: A decade of research. *Ethics & Behavior, 11*, 219-232. doi:10.1207/S15327019EB1103_2
- Meade, A. W., & Craig, S. B. (2012). Identifying careless responses in survey data. *Psychological Methods, 17*, 437-455. doi:10.1037/a0028085

Milliken, R. (1993, February 20). 'Thalidomide doctor' guilty of medical fraud: William McBride, who exposed the danger of one anti-nausea drug, has been disgraced by experiments with another, writes Robert Milliken in Sydney. *The Independent*. Retrieved from <http://www.independent.co.uk/news/world/thalidomide-doctor-guilty-of-medical-fraud-william-mcbride-who-exposed-the-danger-of-one-antinausea-drug-has-been-disgraced-by-experiments-with-another-writes-robert-milliken-in-sydney-1474190.html>

National Health and Medical Research Council, Australian Research Council, & Australian Vice-Chancellors' Committee. (2007). *National statement on ethical conduct in human research 2007 (Updated May 2013)*. Canberra, Australia: Author. Retrieved from <http://www.nhmrc.gov.au/guidelines/publications/e72>

Nonis, S., & Swift, C. O. (2001). An examination of the relationship between academic dishonesty and workplace dishonesty: A multicampus investigation. *Journal of Education for Business, 77*, 69-77. doi:10.1080/08832320109599052

Office of Science and Technology Policy. (2000). *Federal research misconduct policy*. Washington, DC: Author. Retrieved from <http://ori.hhs.gov/federal-research-misconduct-policy>

Pullen, R., Ortloff, V., Casey, S., & Payne, J. B. (2000). Analysis of academic misconduct using unobtrusive research: A study of discarded cheat sheets. *College Student Journal, 34*, 616-625.

Schooler, J. (2011). Unpublished results hide the decline effect. *Nature, 470*, 437. doi:10.1038/470437a

- Steen, R. G. (2012). Retractions in the scientific literature: Is the incidence of research fraud increasing? *Journal of Medical Ethics, 37*, 249-253. doi:10.1136/jme.2010.040923
- Steneck, N. H. (2006). Fostering integrity in research: Definitions, current knowledge and future directions. *Science and Engineering Ethics, 12*, 53-74. doi:10.1007/PL00022268
- Stroebe, W., Postmes, T., & Spears, R. (2012). Scientific misconduct and the myth of self-correction in science. *Perspectives on Psychological Science, 7*, 670-688. doi:10.1177/1745691612460687
- Tourangeau, R., & Yan, T. (2007). Sensitive questions in surveys. *Psychological Bulletin, 133*, 859-883. doi:10.1037/0033-2909.133.5.859
- Ward, D. A., & Beck, W. L. (1990). Gender and dishonesty. *The Journal of Social Psychology, 130*, 333-339. doi:10.1080/00224545.1990.9924589
- Wasserman, R. (2013). Ethical issues and guidelines for conducting data analysis in psychological research. *Ethics & Behavior, 23*, 3-15.

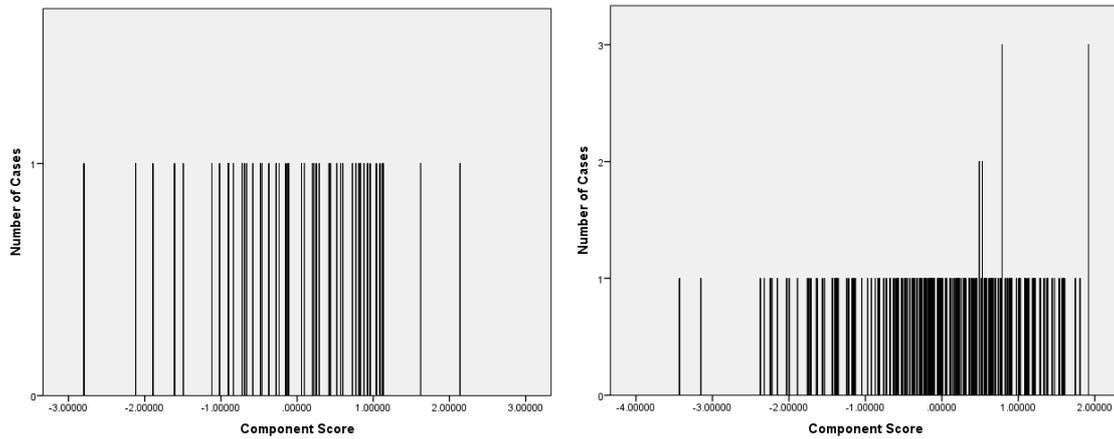


Figure 1. Bar graphs illustrating the absence (left) and presence (right) of cases with identical component scores. Cases with identical component scores have an identical string of responses to the set of variables analysed. There are 10 PPCs visible in the right panel.

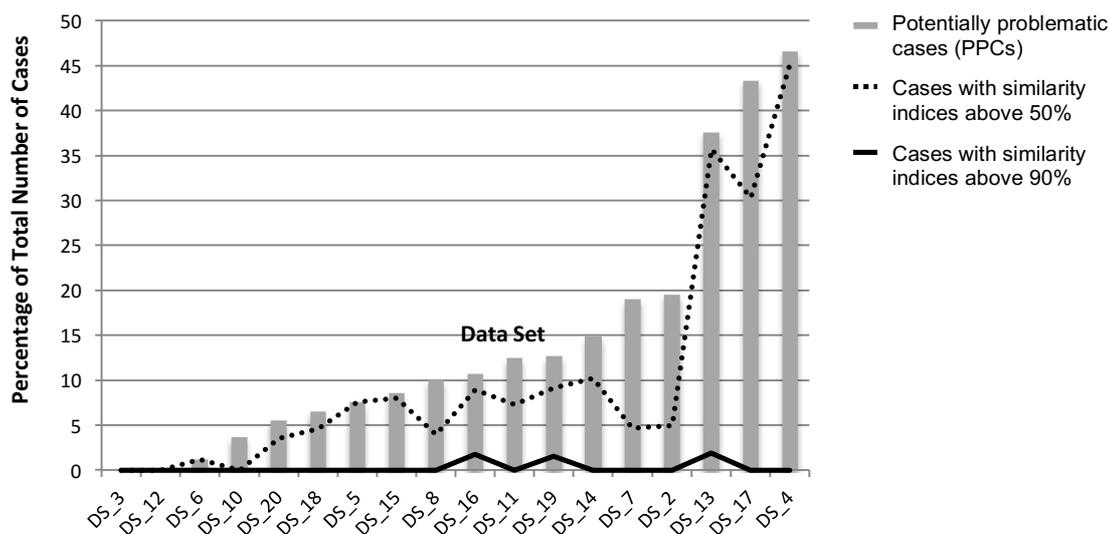


Figure 2. Bar graph illustrating the percentage of PPCs in each data set, along with the percentage of cases in each set with similarity indices above 50% and 90%. Percentages reported relate to the full data sets, as provided by the student researchers. In eight instances, the student researchers reported analyses derived from a reduced data set. However, the only reason cited for excluding cases from analyses was missing data, and component scores were not computed for cases with missing data. Therefore, in all instances, re-running the analyses on the reduced data sets would have inflated the figures reported above.

**Paper 5: The Impact of Academic Sponsorship on Web Survey Dropout and
Item Non-Response**

Allen, P. J., & Roberts, L.D. (2016). The impact of academic sponsorship on Web survey dropout and item non-response. *First Monday*, 21(2). doi:10.5210/fm.v21i2.6144



The impact of academic sponsorship on Web survey dropout and item non-response

by Peter James Allen and Lynne D. Roberts

Abstract

This paper reports two experiments in which the prominence of university sponsorship on Web surveys was systematically manipulated, and its effects on dropout and item non-response were observed. In Study 1, 498 participants were randomised to online surveys with either high or low university sponsorship. Overall, 13.9 percent of participants commenced, but did not complete the surveys, and there was no difference between the proportions of participants dropping out of each condition. However, counter to our predictions, participants in the high sponsorship condition displayed significantly higher item non-response. In Study 2 ($N = 159$), which addressed a rival explanation for the findings in Study 1, the overall dropout rate was 23.9 percent and sponsorship prominence had no effect on either outcome variable. Overall, these findings suggest that hosting information pages on university Web sites, placing university logos on survey pages, and including the name of the university in survey URLs do not reliably impact on dropout or item non-response. Although it may seem disappointing that enhancing sponsor visibility is not sufficient to reduce dropout and item non-response, researchers without ready access to university Web servers or branding will appreciate these findings, as they indicate that minimally visible sponsorship does not necessarily compromise data quality.

Contents

[Introduction](#)

[Study 1](#)

[Study 2](#)

[Discussion](#)

[Conclusion](#)

Introduction

Since the 1993 public release of the first major graphical interface for the World Wide Web, the Mosaic browser, global Internet penetration has increased rapidly (Zakon, 2011). By the end of 2011, around one-third of the world's population was defined by the Internet Telecommunications Union (ITU, 2011a) as "Internet users", although access remains heavily skewed in favour of developed nations (with penetration exceeding 90 percent in parts of Europe; ITU, 2011b), the wealthy, educated and young (Australian Bureau of Statistics [ABS], 2014). In Australia, where the current research was conducted, around 79 percent of the population have regular access to the Internet, mostly at broadband speeds (ABS, 2011). The situation is similar in the U.K. and the U.S. (ITU, 2011b).

As the role of the Internet in everyday life has increased, researchers have sought to exploit the opportunities it affords for data collection (Skitka and Sargis, 2006; Reips, 2007; Lee, *et al.*, 2008). Although a wide variety of different types of research are now conducted either partially or completely online (including qualitative, non-reactive and experimental studies), Web surveying is currently dominant (Reips, 2008; Buchanan and Hvizdak, 2009; Krantz and Williams, 2010), and is continuing to grow in popularity (Lee, *et al.*, 2008).

The popularity of Web surveying can be linked to the many advantages it provides over telephone or paper-based surveying. These include the ability to rapidly access large samples (Skitka and Sargis, 2006; Rentfrow, *et al.*, 2008), which are often more diverse and 'representative' than traditional samples (Gosling, *et al.*, 2004; Lewis, *et al.*, 2009); the ability to connect with rare, geographically disparate or otherwise difficult to access participants (*e.g.*, Hildebrandt and colleagues, 2006, large sample of anabolic steroid users); reduced social desirability and experimenter expectancy effects (Hewson and Laurent, 2008); and, the ability to easily randomize and impose conditional logic on the presentation of survey items and stimuli (Best and Krueger, 2004).

Despite these advantages, there are also a number of challenges associated with Web surveying. For example, researchers cannot easily exert control over the conditions under which participants complete Web surveys, and consequently it's difficult to know if and how divided their attention is during completion (Stieger and Reips, 2010). There are also unique ethical considerations (Allen and Roberts, 2010; Buchanan and Williams, 2010; Roberts and Allen, 2015); concerns about multiple submissions (Reips, 2002); relatively low response rates (*e.g.*, 10–11 percent lower than other surveying methods in two recent meta-analyses; Lozar Manfreda, *et al.*, 2008; Shih and Fan, 2008); higher levels of item non-response (*i.e.*, missing data; Heerwegh and Loosveldt, 2008; Scott, *et al.*, 2011, but see Denscombe, 2009; Dillman, *et al.*, 2010) and relatively high dropout rates (Peytchev, 2009; Rossman, *et al.*, 2011). It is these latter two concerns — item non-response and dropout — that are the focus of the current research.

Dropout and item non-response

Dropout (also referred to as break-off or non-completion) rate can be defined as the proportion of participants who start, but do not finish a Web survey (Heerwegh and Loosveldt, 2006; Ekman, *et al.*, 2007). It is the inverse of retention rate, which is the proportion of participants who reach and complete the final page of a survey (Görizt, 2006b). In a face-to-face or telephone setting, social pressures can inhibit a participant's desire to say 'I want to stop now', regardless of a researcher's assurances that they can withdraw at any time (Buchanan and Williams, 2010). No such pressures exist in an online context, and consequently, drop out rates are often quite high. For example, in the 20 Web experiments described by respondents in Musch and Reips' (2000) survey of online researchers, the mean dropout rate was 34 percent, and ranged from one percent to 87 percent. In a methodologically similar study of Web surveys, dropout rates ranged from 0 percent to 73 percent, with a mean of 16 percent ($N = 68$; Lozar Manfreda and Vehovar, 2002).

Research has identified many factors that either cause, or can predict dropout from Web surveys. Causal factors, which have captured most research attention thus far, include the provision of incentives (Görizt, 2010, 2006a, 2006b; Sauermann and Roach, 2013); the stated and actual length of the survey (Galesic and Bosnjak, 2009; Hoerger, 2010; Yan, *et al.*, 2011) and the burden it places on participants (Crawford, *et al.*, 2001); the use of individual invitations versus general requests for participation during recruitment (Lozar Manfreda and Vehovar, 2002; Heerwegh and Loosveldt, 2006; Sánchez-Fernández, *et al.*, 2012); if and how progress indicators are used (Matzat, *et al.*, 2009; Conrad, *et al.*, 2010; Yan, *et al.*, 2011); the use of the forced-response feature available in most Web surveying applications (Fuchs, 2003; Heerwegh, 2005; Stieger, *et al.*, 2007); how the survey is structured (*e.g.*, one versus many items per page; Lusinci, 2007); and how items are presented and ordered (Heerwegh and Loosveldt, 2002; O'Neil, *et al.*, 2003; Ekman, *et al.*, 2007). Individual differences factors correlated with dropout include the education level of participants (Ekman, *et al.*, 2007; Peytchev, 2009), their student status (O'Neil, *et al.*, 2003; O'Neil and Penrod, 2001) and level of interest in the survey topic

(Roßmann, *et al.*, 2011).

Item non-response occurs when participants do not answer survey questions they have been exposed to, and are eligible to complete (Bosnjak and Tuten, 2001). For practical purposes, response options like "don't know" and "prefer not to say" are typically also treated as item non-response by most researchers, even though it's recognised that these are not perfectly equivalent types of missing data (Albaum, *et al.*, 2011). Research comparing the extent of item non-response across different surveying modes has produced mixed findings. For example, Heerwegh and Loosveldt (2008) found that both item non-response and endorsement of the "don't know" response options were significantly higher for Web survey respondents than face-to-face respondents. However, their effects were small and can be partially attributable to the absence of a "don't know" option on their response cards (as is common practice in face-to-face surveying) coupled with the interviewers' use of probing techniques to elicit responses from participants. When researchers have compared Web to mail-surveys in both experimental (Kwak and Radler, 2002; Bech and Kristensen, 2009; Messer, *et al.*, 2012) and quasi-experimental designs (Haraldsen, *et al.*, 2002; Denscombe, 2006; Lorenc, 2010; Israel and Lamm, 2012; Lesser, *et al.*, 2012), they have tended to find less item non-response in the Web mode. However, this finding is not unequivocal, with Millar and Dillman (2012) and Wolfe, *et al.* (2009) both reporting no differences between modes. When a modal difference is observed, it is typically small for fixed-choice items, but larger for open-ended items (Huang, 2006; Denscombe, 2009), although again this is not always the case (Millar and Dillman, 2012).

Beyond the type of item (*i.e.*, fixed-choice vs. open-ended), presenting items all-at-once rather than one-at-a-time (Nosek, *et al.*, 2012) and including a "no answer" response option in items utilizing drop-down menus (Heerwegh and Loosveldt, 2002) have been found to increase item non-response. Other item types associated with higher non-response include those following branching questions, multiple part, demographic and other "sensitive" items (Messer, *et al.*, 2012; Kays, *et al.*, 2012). Finally, individual differences factors correlated with item non-response include education level, age, income and student status (Dillman, *et al.*, 2010; Messer, *et al.*, 2012). These relationships are typically quite small.

One intuitively appealing remedy for item non-response is the use of the forced-response feature that is available in most online surveying applications. When deployed, this feature prevents a respondent from continuing to the next item or page until the current one has been completed. However, using it often results in significantly higher dropout. For example, Stieger and colleagues (2007) displayed a 'hard prompt' error message each time a respondent attempted to skip a survey item, and found that those exposed to the message were three times more likely to drop out than those who were not. Fuchs (2003) and Heerwegh (2005) reported similar trends, although they did not reach statistical significance in the latter's research, where the overall level of dropout was also much lower.

Because of the undesirable consequences associated with the forced-response feature, we need to consider other techniques and strategies that can be employed to reduce item non-response and dropout. One such potential strategy involves enhancing the prominence of survey sponsorship, which is the independent variable (IV) in the current research.

Survey sponsorship

The sponsor of a survey is the agency or organisation responsible for "funding part or all of the sampling and data collection activities and typically has first or exclusive rights to the data" [1]. Research has indicated that offline surveys with university or government sponsorship tend to yield higher response rates than those sponsored by commercial entities (Heberlein and Baumgartner, 1978; Fox, *et al.*, 1988; Edwards, *et al.*, 2002; Groves and Peytcheva, 2008; but see Yammarino, *et al.*, 1991). It has been argued that this effect can be attributed to the higher prestige, moral authority or legitimacy that university or government sponsorship tends to convey (Groves, *et al.*, 1992; Boulianne, 2008). Furthermore, compared to surveys by commercial organisations, those by university or government departments are less likely to be sales calls disguised as 'research', and more likely to contribute to the advancement of science or well-being of the community (Boulianne, 2008). Although no research has directly examined the effects of sponsor type on Web survey response rates, intent to respond to a Web survey has been predicted by trust in its sponsor (Fang, *et al.*, 2009) as well as the survey sponsor's reputation (Fang, *et al.*, 2012). Finally, when Boulianne, *et al.* (2011) manipulated Web survey sponsor prominence, such that members of a university community were invited to complete a survey about transportation issues by either the university's transportation department or its survey centre, it had no impact on response rate. The absence of any effect in Boulianne and colleagues' study is perhaps unsurprising, considering the subtlety of their manipulation, in which both survey invitations came from different departments within the same university, and the invitation from the survey centre clearly indicated that the transportation department was actually conducting the research.

Research examining the effects of sponsorship on survey dropout and item non-response is more limited, and the findings are mixed. In an off-line context, Peterson (1975) reported higher item non-response for a business sponsored survey compared to a university-sponsored survey, whereas Jones and Linda (1978) reported no differences in item non-response between business, university and government sponsored surveys. When Etter, *et al.* (1996) compared mail surveys sponsored by either a private medical practice or a university, they similarly found no item non-response differences. Online, it is possible to also study dropout behaviour in relation to survey sponsorship, although few researchers have done so, and again their findings have been inconsistent. For example, both Heerwegh and Loosveldt (2006) and Boulianne and colleagues (2011) manipulated the prominence of Web survey sponsorship, although only the latter observed a reliable effect. However, Heerwegh and Loosveldt's study, the experimental condition in which participants were exposed to the sponsoring university's logo on every page of the survey did exhibit lower dropout, albeit not significantly lower.

Rationale and hypotheses

There is an absence of experimental research examining the effects of Web survey sponsorship on both dropout and item non-response. The online studies that have been conducted have only examined item non-response, and have produced inconsistent findings. Furthermore, the available off-line studies have predominantly manipulated the nature of the survey sponsor, rather than its prominence or intensity. However, in authentic research situations it is far more likely that a researcher will be able to enhance or reduce sponsorship prominence (*e.g.*, by adding or removing sponsor branding) than be able to change the nature of the sponsor (*e.g.*, from a commercial to a non-commercial sponsor) in his or her efforts to reduce dropout and item non-response.

The current research addresses these deficits in the literature, and also tackles a practical question that we have asked over several years as supervisors of undergraduate psychology dissertation research, which has increasingly become survey-based and online at our university. Specifically, we have wanted to know if the extra time and effort involved in securing permission to use corporate university branding on Web surveys and hosting information pages on university Web sites actually have a reliable impact on the quality of Web survey data. To answer this question, we describe two experiments in which the prominence of university sponsorship on Web surveys was systematically manipulated, and its effects on dropout and item non-response were observed.

At a time when online methods are increasingly dominating survey research, it is important that we expend effort on studying easily manipulated variables that have the potential to affect data quality, which has long been a concern to survey researchers (*e.g.*, Blasius and Thiessen, 2012). Dropout and item non-response are two key indicators of data quality. They reduce the overall volume of data available for analysis (with consequent implications for statistical power), impact on the representativeness and generalizability of findings, and can also raise difficult ethical issues (*e.g.*, should a respondent's data be included in analyses if they drop out on the final page of a Web survey? de Leeuw, *et al.*, 2003; Denscombe, 2009; Roßmann, *et al.*, 2011). One such easily manipulated variable is sponsorship prominence and, based on prior empirical and theoretical (*e.g.*, the authority principle; social exchange theory; see Heerwegh and Loosveldt, 2006) work, we hypothesised that, compared to Web surveys displaying a low level of university sponsorship, those displaying a high level would result in reduced (*H1*) dropout and (*H2*) item non-response.

Study 1

Method

Participants

A convenience sample of 498 adults were recruited via face-to-face (*e.g.*, flyers) and electronic (*e.g.*, e-mail messages and links on

Web sites) methods. Amongst those who provided demographic data (over 80 percent of the sample), the mean age was 24.49 years ($SD = 7.89$) and gender was evenly split. Sixty percent of the sample identified as students, while less than six percent identified as either unemployed or retired. The majority (95 percent) of the sample reported that they access the Internet over a broadband connection.

Prior to recruiting participants, this study was reviewed and approved by our local Human Research Ethics Committee (HREC). Participants were treated in accordance with the Australian National Health and Medical Research Council's (2007) statement on ethical conduct in human research. No compensation was provided for participation, however participants were offered the opportunity to enter a prize draw as a token of our appreciation for their time.

A 78-item, 10-page online 'Internet Piracy Survey' was used to collect the data reported in this study. The survey contained measures of privacy concern (16 items; Buchanan, *et al.*, 2007), psychological reactance (11 items; Hong and Faedda, 1996), perceived behavioural control for pirating digital content (7 items; Cronan and Al-Rafee, 2008; Wang, *et al.*, 2009), intent, attitudes and subjective norms regarding piracy (19 items; Cronan and Al-Rafee, 2008), the perceived legitimacy of digital publishing companies (12 items; Wolf, 2009), piracy behaviour (six items), and seven demographic questions. These measures primarily used Likert-type response formats, although several items used check-boxes and text-fields instead. In the current study, we are not concerned with participants' substantive responses to these measures; only whether or not they completed the survey, and the total number of items responded to.

Four versions of the survey were developed that were identical in content, but differed in presentation format. The first pair of surveys were hosted on our faculty Web server using LimeSurvey (<http://limesurvey.org>), and were preceded by an information page on our school Web site. Our university logo featured prominently on every page of these surveys, representing "high" university sponsorship, which is the first level of the IV in this study. The second level of the IV, "low" university sponsorship, was characterised by a pair of information sheets and surveys that were hosted on SurveyMonkey.com. Our university logo did not appear anywhere on these surveys, although its name was mentioned twice in the information sheets, in accordance with institutional ethical requirements. One version of each pair "forced" participants to answer every question on each page before continuing, whereas the other did not (*i.e.*, all questions were "optional"). Both LimeSurvey and SurveyMonkey were selected due to their popularity (Allen and Roberts, 2010) and comparable feature sets.

The information page that preceded each version of the survey described the research as investigating factors influencing Internet piracy and survey completion behaviours, but did not explicitly mention the experimental manipulation. However, at the end of each survey, participants were automatically re-directed to a page on our school Web site that revealed the full nature of the study.

Dropout, the first dependent variable (DV) in this study, was operationalised as whether or not the participant clicked the "submit" button at the end of the survey. Item non-response, the second DV, was operationalised as the number of items (out of 78) that the participant provided a response to.

Procedure

Prospective participants were initially directed to a page on our school Web site, which simply thanked them for their interest in the study and requested that they click on a link to continue to the survey. Attached to this link was a Perl script (Wright, 1996), which automatically randomised each participant to one of the four versions of the survey. The only way to detect the presence of the Perl script on this page was to examine its source code. Participants then read through the information sheet, worked through the 78 items on the relevant version of the survey and, if applicable, were re-directed back to the school Web site on completion. The four survey groups were of a statistically equivalent size, and did not differ on any of the demographic characteristics measured, suggesting that the randomisation was successful.

Results and discussion

Overall, 13.9 percent of participants commenced, but did not complete the surveys. The proportion of participants who *completed* the high sponsorship surveys (.856) did not differ from the proportion who *completed* the low sponsorship surveys (.867), 95 percent CI of the difference between proportions $[-.050, .072]$, $\chi^2(1, N = 498) = 0.13$, $p = .718$, two-tailed, $\phi = .016$.

Of those who *completed* the optional format surveys ($n = 216$, representing 87.10 percent of participants exposed to this format), members of the high sponsorship condition ($Mdn = 77.00$) answered significantly fewer items than members of the low sponsorship condition ($Mdn = 78.00$), Hodges-Lehman 95 percent CI of the median difference $[-1.00, 0.00]$, $U = 3734.50$, $z = -5.29$ (corrected for ties), $p < .001$, two-tailed. This difference could be described as medium-sized, $r = .36$. There was no such difference between members of the high ($Mdn = 21.50$) and low ($Mdn = 27.00$) conditions who *did not complete* the optional response surveys ($n = 32$, representing 12.9 percent of participants exposed to this format), Hodges-Lehman 95 percent CI of the median difference $[-27.00, 1.00]$, $U = 90.50$, $z = -1.17$ (corrected for ties), $p = .243$, two-tailed. However, this difference was non-trivial ($r = .21$), and thus non-significance should be interpreted with caution.

Counter to our predictions, there is some evidence to suggest that reducing the prominence of university sponsorship may increase the number of items that participants respond to in online surveys utilising an optional response format. However, LimeSurvey and SurveyMonkey differ in terms of basic page formatting, load speeds and several other factors, which could be responsible for these findings. These confounds were addressed in Study 2, which delivered both high and low sponsorship surveys on the same survey platform (Qualtrics.com).

Study 2

Method

Participants

A convenience sample of 159 adults were recruited via face-to-face (*e.g.*, flyers) and electronic (*e.g.*, e-mail messages and links on Web sites) methods in mid-2011. Amongst those who provided demographic data (over 90 percent of the sample), the median age range was 21–30 years, and 70 percent were female. They were treated in accordance with local ethical guidelines, and were not offered any incentives or compensation for participation.

Measures

A 65-item, seven-page online 'Internet Behaviour Survey' contained five demographic items as well as measures of Internet use (six items), privacy concern (16 items; Buchanan, *et al.*, 2007), perceived credibility of the survey sponsor (three items; Rifon, *et al.*, 2004), attitudes toward the survey sponsor (three items; Mackenzie and Lutz, 1989), trust in the survey sponsor (three items; Fang, *et al.*, 2009), and willingness to disclose personal information in online surveys (29 items; Joinson, *et al.*, 2008). These measures used a variety of response formats, including Likert-type, semantic differential, check boxes and text fields. Like Study 1, we are not concerned with participants' substantive responses to these measures; only whether or not they completed the survey, and the total number of items responded to.

To operationalise the IV for this study, two versions of the survey were developed that were identical in content, utilised an optional response format, and were hosted on Qualtrics.com. The first survey represented a high level of university sponsorship, was preceded by an information page on our school Web site, and had the university name and logo featured prominently on every page, and in the survey URL. The information sheet for the second survey was hosted on Qualtrics.com, along with the survey itself. There were no university logos displayed on this version of the survey, and the survey URL did not contain the university name. It should be noted however that the university name was mentioned twice in the information sheet, in accordance with institutional ethical requirements.

The information page that preceded each version of the survey described the research as investigating Internet behaviour and factors influencing how people respond to online surveys, but did not explicitly mention the experimental manipulation. However, at the end of each survey, participants were automatically re-directed to a page on our school Web site that revealed the full nature of the study.

Dropout, the first DV in this study, was operationalised as whether or not the participant clicked the "submit" button at the end of the survey. Item non-response, the second DV, was operationalised as the number of items (out of 65) that the participant provided a response to. Note that 26 items offered participants a "prefer not to say" option, which was coded as the absence of a response for the purposes of data analysis.

Procedure

Prospective participants were initially directed to <http://internetbehavioursurvey.com> (no longer live), which simply thanked them for their interest in the study and requested that they click on a link to continue to the survey. Attached to this link was the same Perl script (Wright, 1996) used in Study 1, which automatically randomised each participant to one of the two versions of the survey. Participants then read the information sheet, worked through the 65 items on the relevant version of the survey and, if applicable, were re-directed back to the school Web site on completion. The two survey groups were of a statistically equivalent size, and did not differ on age or gender, suggesting that the randomisation was successful.

Results

Overall, 23.9 percent of participants commenced, but did not complete the surveys. This was a significantly higher dropout rate than observed in Study 1, $\chi^2(1, N = 657) = 8.92, p = .003$, two-tailed, although the effect size was relatively small, $\phi = .117$. The proportion of participants that completed the high sponsorship survey (.763) did not differ from the proportion that completed the low sponsorship survey (.759), 95 percent CI of the difference between proportions [-.135, .129], $\chi^2(1, N = 159) = 0.002, p = .965$, two-tailed, $\phi = .004$. Furthermore, the numbers of items answered by participants who completed the high ($Mdn = 65$ items) and low ($Mdn = 64$ items) sponsorship surveys were statistically equivalent, Hodges-Lehman 95 percent CI of the median difference [-1.00, 0.00], $U = 1662.50, z = -0.92$ (corrected for ties), $p = .356$, two tailed, $r = .084$. A similar pattern of results was found for those participants who did not complete the high ($Mdn = 9$) and low ($Mdn = 9$) sponsorship surveys, Hodges-Lehman 95 percent CI of the median difference [-9.00, 12.00], $U = 171.50, z = -0.27$ (corrected for ties), $p = .789$, two tailed, $r = .043$ [2].

Discussion

In this paper we have described two studies in which the prominence of university sponsorship on Web surveys was systematically manipulated, and its effects on dropout and item non-response were observed. In the first study, findings indicated that a high level of university sponsorship might actually increase item non-response. However, when alternative plausible explanations for this effect were ruled out in Study 2, it disappeared. In neither study did we observe any effect of sponsorship prominence on dropout. Overall, these findings lead us to conclude that hosting information pages on university Web sites, placing university logos on survey pages, and including the name of the university in survey URLs do not reliably impact on dropout or item non-response. However, these measures may provide other benefits, such as enhancing the honesty and candidity of responding, or improving response rates. These issues will require investigation in future research.

On the surface, these findings may seem disappointing, as when viewed in conjunction with Heerwegh and Loosveldt (2006) they suggest that simply enhancing sponsor visibility is not sufficient to reliably reduce dropout and item non-response. However, researchers, and particularly student researchers, without ready access to university Web servers or branding will appreciate these findings, as they indicate that minimally visible sponsorship does not necessarily compromise data quality.

It should be noted that our low sponsorship condition did not reflect a complete absence of sponsorship, as pragmatic and institutional considerations meant that our affiliation with a university would still have been obvious to most participants. For example, invitations to participate were sent from university e-mail addresses, and the information sheet clearly indicated that the study was being conducted by university-based researchers, and had been approved by a university HREC. However, these are minimum standards that all legitimate researchers ought to be able to meet, and thus it could be argued that attempting to reduce the prominence of sponsorship below this level would only decrease the ecological validity of our experimental manipulation.

It should also be noted that this study only investigated manipulating the prominence of sponsorship by one Australian university, which typically ranks towards the bottom of the top third of Australian universities on commonly cited indices of performance (e.g., *Times Higher Education Supplement*, 2012). It says nothing about the prominence of other types of sponsorship, such as that of internationally renowned universities [3], or different types of government or commercial entities. Nor does it speak to possible interactions between sponsor prominence, affiliations with the sponsor, knowledge or it, or attitudes towards it. Again, these are possible topics for future research.

Conclusion

In summary, this paper presents two studies in which the prominence of a university sponsor on a Web survey was systematically varied, and its effects on dropout and item non-response were observed. Their findings lead to the conclusion that the prominence of university sponsorship affects neither, although it is yet unknown whether it affects the initial decision about whether or not to participate in a piece of research, or decisions about how to respond to specific survey items. It is anticipated that future research will shed light on these issues. 

About the authors

Peter J. Allen is Lecturer in the School of Psychology and Speech Pathology at Curtin University in Perth, Western Australia. Direct comments to: [p \[dot\] allen \[at\] curtin \[dot\] edu \[dot\] au](mailto:p[dot]allen[at]curtin[dot]edu[dot]au)

Lynne D. Roberts is an Associate Professor in the School of Psychology and Speech Pathology at Curtin University, and an OLT National Teaching Fellow. E-mail: [lynne \[dot\] roberts \[at\] curtin \[dot\] edu \[dot\] au](mailto:lynne[dot]roberts[at]curtin[dot]edu[dot]au)

Notes

1. Boulianne, 2008, p. 868.

2. For these two Mann-Whitney U tests, "prefer not to say" (which was a response option for 26 of the items on the Joinson, *et al.*, 2008, measure) was coded as the absence of a response. Treating "prefer not to say" as a response yields essentially equivalent results. Interestingly, the proportion of our participants making use of "prefer not to say" was far lower than that observed by Joinson, *et al.* (2008). On average, it was used by less than 2.5 percent of participants on each item, and was selected most frequently for the items about participants' previous sexual partners (12.4 percent of responses), visits to their doctor (9.9 percent) and support for the death penalty (9.1 percent).

3. Although it should be noted that the sponsor in Heerwegh and Loosveldt's (2006) research is the top ranking university in Belgium, and one of the top 20 in Europe, according to the *Times Higher Education Supplement* (2012).

References

- Gerald Albaum, James Wiley, Catherine Roster, and Scott M. Smith, 2011. "Visiting item non-responses in Internet survey data collection," *International Journal of Market Research*, volume 53, number 5, pp. 687–703. doi: <http://dx.doi.org/10.2501/IJMR-53-5-687-703>, accessed 17 January 2016.
- Peter J. Allen and Lynne D. Roberts, 2010. "The ethics of outsourcing online survey research," *International Journal of Technoethics*, volume 1, number 3, pp. 35–48. doi: <http://dx.doi.org/10.4018/jte.2010070104>, accessed 17 January 2016.
- Australian Bureau of Statistics, 2014. "Household use of information technology, Australia, 2012–13" (25 February), at <http://www.abs.gov.au/Ausstats/abs@.nsf/mf/8146.0>, accessed 3 July 2015.
- Mickael Bech and Morten Bo Kristensen, 2009. "Differential response rates in postal and Web-based surveys among older respondents," *Survey Research Methods*, volume 3, number 1, at <https://ojs.ub.uni-konstanz.de/srm/article/view/592>, accessed 17 January 2016. doi: <http://dx.doi.org/10.18148/srm/2009.v3i1.592>, accessed 17 January 2016.
- Samuel J. Best and Brian S. Krueger, 2004. *Internet data collection*. Thousand Oaks, Calif.: Sage.
- Jörg Blasius and Victor Thiessen, 2012. *Assessing the quality of research data*. London: Sage.
- Michael Bosnjak and Tracy L. Tuten, 2001. "Classifying response behaviors in Web-based surveys," *Journal of Computer-Mediated Communication*, volume 6, number 3, at <http://onlinelibrary.wiley.com/doi/10.1111/j.1083-6101.2001.tb00124.x/full>, accessed 3 July 2015. doi: <http://dx.doi.org/10.1111/j.1083-6101.2001.tb00124.x>, accessed 17 January 2016.
- Shelley Boulianne, 2008. "Survey sponsor," in: Paul J. Lavrakas (editor). *Encyclopedia of survey research methods*. Thousand Oaks, Calif.: Sage, pp. 868–869.
- Shelley Boulianne, Casey A. Klofstad, and Danna Basson, 2011. "Sponsor prominence and response patterns to an online survey," *International Journal of Public Opinion Research*, volume 23, number 1, pp. 79–87. doi: <http://dx.doi.org/10.1093/ijpor/edq026>, accessed 17 January 2016.
- Elizabeth A. Buchanan and Erin E. Hvizdak, 2009. "Online survey tools: Ethical and methodological concerns of human research ethics committees," *Journal of Empirical Research on Human Research Ethics*, volume 4, number 2, pp. 37–48. doi: <http://dx.doi.org/10.1525/jer.2009.4.2.37>, accessed 17 January 2016.
- Tom Buchanan and John E. Williams, 2010. "Ethical issues in psychological research on the Internet," In: Samuel D. Gosling and John A. Johnson (editors). *Advanced methods for conducting online behavioral research*. Washington, D.C.: American Psychological Association, pp. 255–271. doi: <http://dx.doi.org/10.1037/12076-016>, accessed 17 January 2016.
- Tom Buchanan, Carina Paine, Adam N. Joinson, and Ulf-Dietrich Reips, 2007. "Development of measures of online privacy concern and protection for use on the Internet," *Journal of the American Society for Information Science and Technology*, volume 58, number 2, pp. 157–165. doi: <http://dx.doi.org/10.1002/asi.20459>, accessed 17 January 2016.
- Scott D. Crawford, Mick P. Couper, and Mark J. Lamias, 2001. "Web surveys: Perceptions of burden," *Social Science Computer Review*, volume 19, number 2, pp. 146–162. doi: <http://dx.doi.org/10.1177/089443930101900202>, accessed 17 January 2016.
- Edith D. de Leeuw, Joop Hox, and Mark Huisman, 2003. "Prevention and treatment of item nonresponse," *Journal of Official Statistics*, volume 19, number 2, pp. 153–176, and at <http://www.ios.nu/Articles/abstract.asp?article=192153>, accessed 17 January 2016.
- Martyn Denscombe, 2009. "Item non-response rates: A comparison of online and paper questionnaires," *International Journal of Social Research Methodology*, volume 12, number 4, pp. 281–291. doi: <http://dx.doi.org/10.1080/13645570802054706>, accessed 17 January 2016.
- Martyn Denscombe, 2006. "Web-based questionnaires and the mode effect: An evaluation based on completion rates and data contents of near-identical questionnaires delivered in different modes," *Social Science Computer Review*, volume 24, number 2, pp. 246–254. doi: <http://dx.doi.org/10.1177/0894439305284522>, accessed 17 January 2016.
- Don A. Dillman, Benjamin L. Messer, Morgan M. Millar, and Michelle L. Edwards, 2010. "Item non-response differences between Web and mail surveys of the general public," paper presented at the International Total Survey Error Workshop (14 June, Stowe, Vt.), at http://www.niss.org/sites/default/files/Dillman%20Stowe_ItemNonresponsePresentation_Dillman_Messer_.pdf, accessed 3 July 2015.
- Phil Edwards, Ian Roberts, Mike Clarke, Carolyn DiGuseppi, Sarah Pratap, Reinhard Wentz, and Irene Kwan, 2002. "Increasing response rates to postal questionnaires: Systematic review," *British Medical Journal*, volume 324, pp. 1,183–1,191. doi: <http://dx.doi.org/10.1136/bmj.324.7347.1183>, accessed 17 January 2016.
- Alexandra Ekman, Åsa Klint, Paul W. Dickman, Hans-Olov Adami, and Jan-Eric Litton, 2007. "Optimizing the design of Web-based questionnaires — Experience from a population-based study among 50,000 women," *European Journal of Epidemiology*, volume 22, number 5, pp. 293–300. doi: <http://dx.doi.org/10.1007/s10654-006-9091-0>, accessed 17 January 2016.
- Jean-François Etter, Thomas V. Perneger, and André Rougemont, 1996. "Does sponsorship matter in patient satisfaction surveys? A randomized trial," *Medical Care*, volume 34, number 4, pp. 327–335.
- Jiaming Fang, Chao Wen, and Robert Pavur, 2012. "Participation willingness in Web Surveys: Exploring effect of sponsoring corporation's and survey provider's reputation," *Cyberpsychology, Behavior, and Social Networking*, volume 15, number 4, pp. 195–199. doi: <http://dx.doi.org/10.1089/cyber.2011.0411>, accessed 17 January 2016.
- Jiaming Fang, Peiji Shao, and George Lan, 2009. "Effects of innovativeness and trust on Web survey participation," *Computers in Human Behavior*, volume 25, number 1, pp. 144–152. doi: <http://dx.doi.org/10.1016/j.chb.2008.08.002>, accessed 17 January 2016.
- Richard J. Fox, Melvin R. Crask, and Jonghoon Kim, 1988. "Mail survey response rate: A meta-analysis of selected techniques for inducing response," *Public Opinion Quarterly*, volume 52, number 4, pp. 467–491. doi: <http://dx.doi.org/10.1086/269125>, accessed 17 January 2016.
- Marek Fuchs, 2003. "Item non-response reminder in a Web survey," paper presented at Methodology and Statistics International Conference (14–17 September, Ljubljana, Slovenia), at <http://vlado.fmf.uni-lj.si/trubar/preddvor/2003/>, accessed 3 July 2015.
- Mirta Galesic and Michael Boznjak, 2009. "Effects of questionnaire length on participation and indicators of response quality in a Web survey," *Public Opinion Quarterly*, volume 73, number 2, pp. 349–360. doi: <http://dx.doi.org/10.1093/poq/nfp031>, accessed 17 January 2016.
- Anja S. Göritz, 2010. "Using lotteries, loyalty points, and other incentives to increase participant response and completion," In: Samuel D. Gosling and John A. Johnson (editors). *Advanced methods for conducting online behavioral research*. Washington, D.C.: American Psychological Association, pp. 219–234. doi: <http://dx.doi.org/10.1037/12076-014>, accessed 17 January 2016.
- Anja S. Göritz, 2006a. "Cash lotteries as incentives in online panels," *Social Science Computer Review*, volume 24, number 4, pp. 445–

459.

doi: <http://dx.doi.org/10.1177/0894439305286127>, accessed 17 January 2016.

Anja S. Göritz, 2006b. "Incentives in Web studies: Methodological issues and a review," *International Journal of Internet Science*, volume 1, number 1, pp. 58–70, at http://www.ijis.net/ijis1_1/ijis1_1_goeertz.pdf, accessed 3 July 2015.

Samuel D. Gosling, Simine Vazire, Sanjay Srivastava, and Oliver P. John, 2004. "Should we trust Web-based studies? A comparative analysis of six preconceptions about Internet questionnaires," *American Psychologist*, volume 59, number 2, pp. 93–104. doi: <http://dx.doi.org/10.1037/0003-066X.59.2.93>, accessed 17 January 2016.

Robert M. Groves and Emilia Peytcheva, 2008. "The impact of nonresponse rates on nonresponse bias: A meta-analysis," *Public Opinion Quarterly*, volume 72, number 2, pp. 167–189. doi: <http://dx.doi.org/10.1093/poq/nfn011>, accessed 17 January 2016.

Robert M. Groves, Robert B. Cialdini, and Mick P. Couper, 1992. "Understanding the decision to participate in a survey," *Public Opinion Quarterly*, volume 56, number 4, pp. 475–495. doi: <http://dx.doi.org/10.1086/269338>, accessed 17 January 2016.

Gustav Haraldsen, Trine Dale, Elisabeth Dalheim, and Halvor Strømme, 2002. "Mode effects in a mail plus Internet designed census," paper presented at the International Conference on Improving Surveys (26–28 August, Copenhagen), at http://www.websm.org/upload/ediator/D2_6_4.pdf, accessed 3 July 2015.

Thomas A. Heberlein and Robert Baumgartner, 1978. "Factors affecting response rates to mailed questionnaires: A quantitative analysis of the published literature," *American Sociological Review*, volume 43, number 4, pp. 447–2462. doi: <http://dx.doi.org/10.2307/2094771>, accessed 17 January 2016.

Dirk Heerwegh, 2005. "Web surveys: Explaining and reducing unit nonresponse, item nonresponse and partial nonresponse," Ph.D. dissertation, Faculteit Sociale Wetenschappen, Katholieke Universiteit Leuven, at https://perswww.kuleuven.be/~u0034437/public/Files/Heerwegh_PhD_Diss_distribute.pdf, accessed 3 July 2015.

Dirk Heerwegh and Geert Loosveldt, 2008. "Face-to-face versus Web surveying in a high-Internet-coverage population: Differences in response quality," *Public Opinion Quarterly*, volume 72, number 5, pp. 836–846. doi: <http://dx.doi.org/10.1093/poq/nfn045>, accessed 17 January 2016.

Dirk Heerwegh and Geert Loosveldt, 2006. "An experimental study on the effects of personalization, survey length statements, progress indicators, and survey sponsor logos in Web surveys," *Journal of Official Statistics*, volume 22, number 2, pp. 191–210, at <http://www.jos.nu/Articles/abstract.asp?article=222191>, accessed 17 January 2016.

Dirk Heerwegh and Geert Loosveldt, 2002. "An evaluation of the effect of response formats on data quality in Web surveys," *Social Science Computer Review*, volume 20, number 4, pp. 471–484. doi: <http://dx.doi.org/10.1177/089443902237323>, accessed 17 January 2016.

Claire Hewson and Dianna Laurent, 2008. "Research design and tools for Internet research," In: Nigel G. Fielding, Raymond M. Lee, and Grant Blank (editors). *Sage handbook of online research methods*. London: Sage, pp 58–78.

Tom Hildebrandt, James Langenbucher, Sasha Carr, Pilar Sanjuan, and Steff Park, 2006. "Predicting intentions for long-term anabolic-androgenic steroid use among men: A covariance structure model," *Psychology of Addictive Behaviors*, volume 20, number 3, pp. 234–240. doi: <http://dx.doi.org/10.1037/0893-164X.20.3.234>, accessed 17 January 2016.

Michael Hoerger, 2010. "Participant dropout as a function of survey length in Internet-mediated university studies: Implications for study design and voluntary participation in psychological research," *Cyberpsychology, Behavior, and Social Networking*, volume 13, number 6, pp. 697–700. doi: <http://dx.doi.org/10.1089/cyber.2009.0445>, accessed 17 January 2016.

Hsiu-Mei Huang, 2006. "Do print and Web surveys provide the same results?" *Computers in Human Behavior*, volume 22, number 3, pp. 334–350. doi: <http://dx.doi.org/10.1016/j.chb.2004.09.012>, accessed 17 January 2016.

International Telecommunications Union, 2011a. "The world in 2011: ICT facts and figures," at <http://www.itu.int/ITU-D/ict/facts/2011/material/ICTFactsFigures2011.pdf>, accessed 3 July 2015.

International Telecommunications Union, 2011b. *ITU world telecommunication/ICT indicators database*. Fifteenth edition. Geneva: International Telecommunications Union; also at <https://www.itu.int/pub/D-IND-WTID-OL>, accessed 17 January 2016.

Glenn D. Israel and Alexa J. Lamm, 2012. "Item nonresponse in a client survey of the general public," *Survey Practice*, volume 5, number 2, at <http://www.surveypractice.org/index.php/SurveyPractice/article/view/47/html>, accessed 3 July 2015.

Adam N. Joinson, Carina Paine, Tom Buchanan, and Ulf-Dietrich Reips, 2008. "Measuring self-disclosure online: Blurring and non-response to sensitive items in Web-based surveys," *Computers in Human Behavior*, volume 24, number 5, pp. 2,158–2,171. doi: <http://dx.doi.org/10.1016/j.chb.2007.10.005>, accessed 17 January 2016.

Wesley H. Jones and Gerald Linda, 1978. "Multiple criteria effects in a mail survey experiment," *Journal of Marketing Research*, volume 15, number 2, pp. 280–284. doi: <http://dx.doi.org/10.2307/3151263>, accessed 17 January 2016.

Kristina Kays, Kathleen Gathercoal, and William Buhrow, 2012. "Does survey format influence self-disclosure on sensitive question items?" *Computers in Human Behavior*, volume 28, number 1, pp. 251–256. doi: <http://dx.doi.org/10.1016/j.chb.2011.09.007>, accessed 17 January 2016.

John H. Krantz and John E. Williams, 2010. "Using graphics, photographs, and dynamic media," In: Samuel D. Gosling and John A. Johnson (editors). *Advanced methods for conducting online behavioral research*. Washington, D.C.: American Psychological Association, pp. 45–61. doi: <http://dx.doi.org/10.1037/12076-004>, accessed 17 January 2016.

Nojin Kwak and Barry Radler, 2002. "Comparison between mail and Web surveys: Response pattern, respondent profile, and data quality," *Journal of Official Statistics*, volume 18, number 2, pp. 257–273, and at <http://www.jos.nu/Articles/abstract.asp?article=182257>, accessed 17 January 2016.

Raymond M. Lee, Nigel Fielding, and Grant Blank, 2008. "The Internet as a research medium: An editorial introduction to the *Sage handbook of online research methods*," In: Nigel G. Fielding, Raymond M. Lee, and Grant Blank (editors). *Sage handbook of online research methods*. London: Sage, pp 3–20.

Virginia M. Lesser, Lydia A. Newton, and Daniel Yang, 2012. "Comparing item nonresponse across different delivery modes in general population surveys," *Survey Practice*, volume 5, number 2, at <http://www.surveypractice.org/index.php/SurveyPractice/article/view/46>, accessed 3 July 2015.

Ioni Lewis, Barry Watson, and Katherine M. White, 2009. "Internet versus paper-and-pencil survey methods in psychological experiments: Equivalence testing of participants' responses to health-related messages," *Australian Journal of Psychology*, volume 61, number 2, pp. 107–116. doi: <http://dx.doi.org/10.1080/00049530802105865>, accessed 17 January 2016.

Boris Lorenc, 2010. "Item nonresponse analysis for a mixed-mode survey," In: Michael Carlson, Hans Nyquist, and Mattias Villani

(editors). *Official statistics — Methodology and applications in honour of Daniel Thorburn*. Stockholm: Department of Statistics, Stockholm University in collaboration with Statistics Sweden, pp 117–136, and at <https://officialstatistics.files.wordpress.com/2010/05/bok11.pdf>, accessed 17 January 2016.

Katja Lozar Manfreda and Vasja Vehovar, 2002. "Survey design features influencing response rates in Web surveys," paper presented at the International Conference on Improving Surveys (ICIS 2002), (25–28 August, Copenhagen), at http://www.websm.org/uploadi/editor/Lozar_Vehovar_2001_Survey_design.pdf, accessed 11 August 2015.

Katja Lozar Manfreda, Michael Boznjak, Jernej Berzelak, Iris Haas, and Vasja Vehovar, 2008. "Web surveys versus other survey modes: A meta-analysis comparing response rates," *International Journal of Market Research*, volume 50, number 1, pp. 79–104.

Dominic Lusinchi, 2007. "Increasing response rates & data quality of Web surveys: Pre-notification and questionnaire paging format," paper presented at the 2007 Federal Committee on Statistical Methodology Research Conference (5–7 November, Arlington, Va.), at http://fcsmsites.usa.gov/files/2014/05/2007FCSM_Lusinchi-X-A.pdf, accessed 11 August 2015.

Uwe Matzat, Chris Snijders, and Wouter van der Horst, 2009. "Effects of different types of progress indicators on drop-out rates in Web surveys," *Social Psychology*, volume 40, issue 1, pp. 43–52.
doi: <http://dx.doi.org/10.1027/1864-9335.40.1.43>, accessed 17 January 2016.

Benjamin L. Messer, Michelle L. Edwards, and Don A. Dillman, 2012. "Determinants of item nonresponse to Web and mail respondents in three address-based mixed-mode surveys of the general public," *Survey Practice*, volume 5, number 2, at <http://www.surveypactice.org/index.php/SurveyPractice/article/view/45/html>, accessed 11 August 2015.

Morgan M. Millar and Don A. Dillman, 2012. "Do mail and Internet surveys produce different item nonresponse rates? An experiment using random mode assignment," *Survey Practice*, volume 5, number 2, at <http://www.surveypactice.org/index.php/SurveyPractice/article/view/48>, accessed 11 August 2015.

Jochen Musch and Ulf-Dietrich Reips, 2000. "A brief history of Web experimenting," In: Michael H. Birnbaum (editor). *Psychological experiments on the Internet*. San Diego, Calif.: Academic Press, pp. 61–87.
doi: <http://dx.doi.org/10.1016/B978-012099980-4/50004-6>, accessed 17 January 2016.

Brian A. Nosek, N. Sriram, and Emily Umansky, 2012. "Presenting survey items one at a time compared to all at once decreases missing data without sacrificing validity in research with Internet volunteers," *PLoS ONE*, volume 7, number 5, e36771 (17 May).
doi: <http://dx.doi.org/10.1371/journal.pone.0036771>, accessed 17 January 2016.

Kevin M. O'Neil and Steven D. Penrod, 2001. "Methodological variables in Web-based research that may affect results: Sample type, monetary incentives, and personal information," *Behavior Research Methods, Instruments, & Computers*, volume 33, number 2, pp. 226–233.
doi: <http://dx.doi.org/10.3758/BF03195369>, accessed 17 January 2016.

Kevin M. O'Neil, Steven D. Penrod, and Brian H. Bornstein, 2003. "Web-based research: Methodological variables' effects on dropout and sample characteristics," *Behavior Research Methods, Instruments, & Computers*, volume 35, number 2, pp. 217–226.
doi: <http://dx.doi.org/10.3758/BF03202544>, accessed 17 January 2016.

Robert A. Peterson, 1975. "An experimental investigation of mail-survey responses," *Journal of Business Research*, volume 3, number 3, pp. 199–210.
doi: [http://dx.doi.org/10.1016/0148-2963\(75\)90022-3](http://dx.doi.org/10.1016/0148-2963(75)90022-3), accessed 17 January 2016.

Andy Peytchev, 2009. "Survey breakoff," *Public Opinion Quarterly*, volume 73, number 1, pp. 74–97.
doi: <http://dx.doi.org/10.1093/poq/nfp014>, accessed 17 January 2016.

Ulf-Dietrich Reips, 2008. "How Internet-mediated research changes science," In: Azy Barak (editor). *Psychological aspects of cyberspace: Theory, research, applications*. Cambridge: Cambridge University Press, pp. 268–294.
doi: <http://dx.doi.org/10.1017/CBO9780511813740.013>, accessed 17 January 2016.

Ulf-Dietrich Reips, 2007. "The methodology of Internet-based experiments," In: Adam Joinson, Katelyn McKenna, Tom Postmes, and Ulf-Dietrich Reips (editors). *Oxford handbook of Internet psychology*. New York: Oxford University Press, pp. 373–390.
doi: <http://dx.doi.org/10.1093/oxfordhb/9780199561803.013.0024>, accessed 17 January 2016.

Ulf-Dietrich Reips, 2002. "Standards for Internet-based experimenting," *Experimental Psychology*, volume 49, number 4, pp. 243–256.
doi: <http://dx.doi.org/10.1026//1618-3169.49.4.243>, accessed 17 January 2016.

Peter J. Rentfrow, Samuel D. Gosling, and Jeff Potter, 2008. "A theory of the emergence, persistence, and expression of geographic variation in psychological characteristics," *Perspectives on Psychological Science*, volume 3, number 5, pp. 339–369.
doi: <http://dx.doi.org/10.1111/j.1745-6924.2008.00084.x>, accessed 17 January 2016.

Lynne D. Roberts and Peter J. Allen, 2015. "Exploring ethical issues associated with using online surveys in educational research," *Educational Research and Evaluation*, volume 21, number 2, pp. 95–108.
doi: <http://dx.doi.org/10.1080/13803611.2015.1024421>, accessed 17 January 2016.

Joss Roßmann, Jan E. Blumenstiel, and Markus Steinbrecher, 2011. "A new approach to the analysis of survey drop-out. Results from a follow-up survey in the German longitudinal election study (GLES)," paper presented at the General Online Research Conference (14–16 March, Dusseldorf), at http://www.websm.org/uploadi/editor/1319895749Rossmann_Blumenstiel_Steinbrecher_2011_A_new_approach_to_the_analysis_of_survey_drop_out_Results_from_Follow-up_Surveys.pdf, accessed 11 August 2015.

Juan Sánchez-Fernández, Francisco Muñoz-Leiva, and Francisco J. Montoro-Ríos, 2012. "Improving retention rate and response quality in Web-based surveys," *Computers in Human Behavior*, volume 28, number 2, pp. 507–514.
doi: <http://dx.doi.org/10.1016/j.chb.2011.10.023>, accessed 17 January 2016.

Henry Sauermann and Michael Roach, 2013. "Increasing Web survey response rates in innovation research: An experimental study of static and dynamic contact design features," *Research Policy*, volume 42, number 1, pp. 273–286.
doi: <http://dx.doi.org/10.1016/j.respol.2012.05.003>, accessed 17 January 2016.

Anthony Scott, Sung-Hee Jeon, Catherine M. Joyce, John S. Humphreys, Guyonne Kalb, Julia Witt, and Anne Leahy, 2011. "A randomised trial and economic evaluation of the effect of response mode on response rate, response bias, and item non-response in a survey of doctors," *BMC Medical Research Methodology*, volume 11, pp. 126–138.
doi: <http://dx.doi.org/10.1186/1471-2288-11-126>, accessed 17 January 2016.

Tse-Hua Shih and Xitao Fan, 2008. "Comparing response rates from Web and mail surveys: A meta-analysis," *Field Methods*, volume 20, number 3, pp. 249–271.
doi: <http://dx.doi.org/10.1177/1525822X08317085>, accessed 17 January 2016.

Linda J. Skitka and Edward G. Sargis, 2006. "The Internet as psychological laboratory," *Annual Review of Psychology*, volume 57, pp. 529–555.
doi: <http://dx.doi.org/10.1146/annurev.psych.57.102904.190048>, accessed 17 January 2016.

Stefan Stieger and Ulf-Dietrich Reips, 2010. "What are participants doing while filling in an online questionnaire: A paradata collection tool and an empirical study," *Computers in Human Behavior*, volume 26, number 6, pp. 1,488–1,495.
doi: <http://dx.doi.org/10.1016/j.chb.2010.05.013>, accessed 17 January 2016.

Stefan Stieger, Ulf-Dietrich Reips, and Martin Voracek, 2007. "Forced-response in online surveys: Bias from reactance and an increase in sex-specific dropout," *Journal of the American Society for Information Science and Technology*, volume 58, number 11, pp. 1,653–

1,660.

doi: <http://dx.doi.org/10.1002/asi.20651>, accessed 17 January 2016.

Times Higher Education Supplement, 2012. "World university rankings 2011–12," at <https://www.timeshighereducation.com/world-university-rankings/2012/world-ranking>, accessed 11 August 2015.

Edward W. Wolfe, Patrick D. Converse, Osaro Airen, and Nancy Bodenhorn, 2009. "Unit and item nonresponses and ancillary information in Web- and paper-based questionnaires administered to school counselors," *Measurement and Evaluation in Counseling and Development*, volume 42, number 2, pp. 92–103.

doi: <http://dx.doi.org/10.1177/0748175609336862>, accessed 17 January 2016.

Matt Wright, 1996. "Matt's script archive: Random link generator," version 1.0 (30 July), at http://www.scriptarchive.com/rand_link.html, accessed 11 August 2015.

Francis J. Yammarino, Steven J. Skinner, and Terry L. Childers, 1991. "Understanding mail survey response behavior: A meta-analysis," *Public Opinion Quarterly*, volume 55, number 4, pp. 613–639.

doi: <http://dx.doi.org/10.1086/269284>, accessed 17 January 2016.

Ting Yan, Frederick G. Conrad, Roger Tourangeau, and Mick P. Couper, 2011. "Should I stay or should I go: The effects of progress feedback, promised task duration, and length of questionnaire on completing Web surveys," *International Journal of Public Opinion Research*, volume 23, number 2, pp. 131–147.

doi: <http://dx.doi.org/10.1093/ijpor/edq046>, accessed 17 January 2016.

Robert H. Zakon, 2011, "Hobbes' Internet timeline," version 10.2, at <http://www.zakon.org/robert/internet/timeline/>, accessed 11 August 2015.

Editorial history

Received 16 July 2015; revised 19 January 2016; accepted 19 January 2016.



"The impact of academic sponsorship on Web survey dropout and item non-response" by Peter J Allen and Lynne D Roberts is licensed under a [Creative Commons Attribution-NonCommercial 4.0 International License](https://creativecommons.org/licenses/by-nc/4.0/).

The impact of academic sponsorship on Web survey dropout and item non-response
by Peter James Allen and Lynne D. Roberts.

First Monday, Volume 21, Number 2 - 1 February 2016

<http://journals.uic.edu/ojs/index.php/fm/rt/prINTERfriendly/6144/5198>

doi: <http://dx.doi.org/10.5210/fm.v21i2.6144>

Paper 6: The Ethics of Outsourcing Online Survey Research

Allen, P. J., & Roberts, L. D. (2010). The ethics of outsourcing online survey research.
International Journal of Technoethics, 1, 35-48. doi:10.4018/jte.2010070104

This paper appears in *International Journal of Technoethics* edited by Rocci Luppicini, Copyright 2010, IGI Global, www.igi-global.com. Reproduced by permission of the publisher.

The Ethics of Outsourcing Online Survey Research

Peter J. Allen, Curtin University, Australia

Lynne D. Roberts, Curtin University, Australia

ABSTRACT

The increasing level of Internet penetration over the last decade has made web surveying a viable option for data collection in academic research. Software tools and services have been developed to facilitate the development and deployment of web surveys. Many academics and research students are outsourcing the design and/or hosting of their web surveys to external service providers, yet ethical issues associated with this use have received limited attention in academic literature. In this article, the authors focus on specific ethical concerns associated with the outsourcing of web surveys with particular reference to external commercial web survey service providers. These include threats to confidentiality and anonymity, the potential for loss of control over decisions about research data, and the reduced credibility of research. Suggested guidelines for academic institutions and researchers in relation to outsourcing aspects of web-based survey research are provided.

Keywords: Anonymity, Confidentiality, Cyberethics, Data Protection, Outsourcing Surveys, SurveyMonkey, Web Survey

INTRODUCTION

Recent Pew Internet and American Life survey data indicate almost three quarters of American adults regularly access the Internet from home (Horrigan, 2009). The vast majority of these connections are at broadband speeds. Data from the Australian Bureau of Statistics (2008), the UK Office for National Statistics (2009) and the OECD's Directorate for Science, Technology and Industry (2009) reveal that Internet penetration levels are similarly high in Australia, the UK, and many other industrialised nations.

As Internet penetration has risen, researchers have increasingly moved their data

collection efforts 'online' (Lee, Fielding, & Blank, 2008; Reips, 2007; Skitka & Sargis, 2006). These efforts have variously involved online interviewing (Hewson, 2007; O'Connor, Madge, Shaw, & Wellens, 2008), observation and other non-reactive methods (Janetzko, 2008; Robinson, 2001), experimentation (Birnbaum, 2007; Reips, 2007) and web surveying (Best & Krueger, 2008; Reips, 2008). Of these online data collection methods, web surveying is currently dominant (Reips, 2008), is continuing to grow in popularity (Lee et al., 2008)¹, is the online method most frequently reviewed by Human Research Ethics Committees (HRECs; Buchanan & Hvizdak, 2009) and thus is the primary focus of this paper.

DOI: 10.4018/jte.2010070104

The growing use of web surveying merits attention to the possible impacts of the technology on research participants. Such ethical considerations are situated within the emerging scholarship on technoethics. Technoethics provides a focus on the ethical considerations associated with technological change (Lupplicini, 2009). Within the broad field of technoethics, Internet ethics and cyber ethics have been identified as key areas (Lupplicini, 2009) with major questions including “What are the ethical responsibilities of Internet researchers to research participants?” (p. 10) and “What are the ethical responsibilities of Internet researchers to protect the identity and confidentiality of data derived from the Internet?” (p. 10). We begin this article by providing an overview of web surveying, including the tools and services that have emerged to facilitate the development and deployment of web surveys. We provide evidence to suggest that commercial web survey hosts are widely used by academic researchers, yet the ethical issues associated with this use have received only limited attention in the academic literature. The main body of this article provides a focus on specific ethical concerns associated with outsourcing aspects of the web surveying process, with particular reference to external commercial web survey hosts. These include threats to confidentiality and anonymity associated with breaches of data protection and the potential loss of control over decisions about the data. Further, the possible impact of externally hosting academic surveys on response rates and responding is examined in terms of online privacy concern and the perceived credibility of research. This article concludes with some suggested guidelines for institutions and researchers in relation to the outsourcing of aspects of academic research utilising web surveys.

Web Surveying

Web surveying typically involves administering a series of questionnaire items of varying types (e.g., rating scales, fixed-choice, open-ended etc.) over the world-wide-web, and can offer a

number of advantages over paper and telephone based surveying methods. Such advantages include, but are not limited to, timely access to large samples (Skitka & Sargis, 2006) that are often more diverse and ‘representative’ than traditional samples (Gosling, Vazire, Srivastava, & John, 2004); access to samples that would otherwise be prohibitively costly or difficult to achieve (e.g., Hildebrandt, Langenbucher, Carr, Sanjuan, & Park’s, 2006) large sample of anabolic steroid users); reduced social desirability and experimenter expectancy effects (Hewson & Laurent, 2008); and the ability to easily randomize and impose conditional logic on the presentation of survey items and stimuli (Best & Krueger, 2004).

The topics that have been investigated using web surveying are diverse, and a full review is beyond the scope of this paper. However, a small sample might include studies typical of Skitka and Sargis’s (2006) three broad categories of web-based research: translational, phenomenological and novel.

Translational studies are those that investigate traditional topics using methods and measures developed offline, and adapted for use on the web. Such adaptation is primarily to capitalize on the efficiencies and global reach afforded by the web. For example, Oliver John, Sam Gosling and colleagues have used online variants of the Big Five Inventory (BFI; see John, Naumann, & Soto, 2008) to collect large volumes of self-report questionnaire data used in a series of investigations into the psychometric properties of the measure, as well as the characteristics and correlates of the ‘big five’ personality factors more broadly (e.g., Rentfrow, Gosling, & Potter, 2008; Robins, Tracy, Trzesniewski, Gosling, & Potter, 2002; Robins, Tracy, Trzesniewski, Potter, & Gosling, 2001; Soto, John, Gosling, & Potter, 2008; Srivastava, John, Gosling, & Potter, 2003). Sample sizes in these studies have ranged from 100,000 to over 600,000 participants (in the case of Rentfrow et al., 2008). Many additional examples of ongoing translational survey research are indexed on websites like Hanover College’s *Psychological Research on the Net*² and the *Web Survey List*³, hosted at the University of Zurich.

Skitka and Sargis's (2006) second category of web-based research, phenomenological, is also well represented on *Psychological Research on the Net* and the *Web Survey List*. Phenomenological web-based research is focused on the nature of Internet behavior itself, and includes examples such as McFarlane, Bull, and Rietmeijer's (2002) study of young adults' online sex seeking behavior, as well as various investigations into 'Internet addiction' (e.g., Greenfield, 1999; Whang, Lee, & Chang, 2003).

Finally, Skitka and Sargis (2006) identified a third category of web-based research, which they referred to as novel. Novel web-based research capitalizes on unique features of the Internet to ask questions that would be methodologically difficult, if not impossible, to address offline. As an example of novel web-based research employing survey methods, Skitka and Sargis cite Vazire and Gosling (2004), who examined the nature and accuracy of personality impressions derived from viewing personal websites.

Web Surveying Tools and Hosting

As the popularity of web surveying has increased, many software tools have been built to facilitate their development and deployment (Kaczmirek, 2008). These tools typically reduce (and often completely eliminate) the specialised programming knowledge that researchers would otherwise require to create and maintain a custom-built online surveying instrument, and can vary greatly in terms of their feature sets, flexibility, usability and cost to the end-user. These tools also vary in the extent to which they require the researcher to outsource aspects of the research (e.g., survey hosting, data collection, storage etc.) to an external service provider.

*SurveyMonkey.com Corporation*⁴ (hereafter *SurveyMonkey*) is one such service provider. It is a commercial venture that provides subscribers with access to a proprietary, browser-based survey editor, which can be used to build and deploy surveys containing a common range of question types (e.g., fixed-choice, open-ended etc.). Surveys constructed

with the *SurveyMonkey* editor, as well as the data they are used to collect, are hosted on the company's secure web-servers. In other words, researchers using *SurveyMonkey* are essentially outsourcing survey formatting, data collection and storage (at least in the short term) to the company.

Although it is a current market leader, *SurveyMonkey* is but one of literally dozens (and probably hundreds) of companies to which survey hosting and data storage can be outsourced. For more exhaustive reviews and evaluations of some of the available alternatives, the reader is directed to Crawford (2002), Beiderniki and Kerschbaumer (2007), Gordon (2002), Wright (2005), Sue and Ritter (2007) and Gaiser and Schreiner (2009). Gaiser and Schreiner, in particular, provide useful guidelines for evaluating commercial web survey hosts based on costs, ease of use, output viewing options and technical support. Many of the more popular outsourcing options are also indexed in the University of Ljubljana's *WebSM*⁵ resource, where they are referred to as "hosted solutions".

Rather than outsourcing, many researchers prefer to, are required to, and/or have the facilities to, host web surveys internally, or 'in-house'. In other words, to host them on web-servers owned and/or managed by the researcher's home institution. In some instances, these surveys will be hand-coded by or for the researcher; in others, they will be developed using standard web authoring software (e.g., *Adobe Dreamweaver*⁶, *Microsoft Expression Web*⁷ etc.), or more specialised survey development applications like *Opinio*⁸ and *Lime Survey*⁹.

Lime Survey is an example of a widely used open-source web application that can be installed on any web-server running MySQL and PHP. *Lime Survey* surveys and databases are typically hosted on the installation web-server. Like *SurveyMonkey*, *Lime Survey* can be used to build and deploy surveys containing a common range of question types. Unlike users of *SurveyMonkey* (and users of closed-source applications such as *Opinio*) users of *Lime Survey* are free to modify and add to its current feature set, a practice that is encouraged amongst open-source

software developers. For a more comprehensive review of open-source surveying options, the reader is referred to Baker (2007). On *WebSM*, both closed- and open-source web surveying applications suitable for building and hosting surveys in-house can be located by browsing for software that runs “on user’s server”.

Universities vary in both the types of software used to develop web surveys, whether surveys are hosted internally or externally, and the policies and procedures surrounding their use. For example, at our institution, Curtin University, both *SurveyMonkey* and *Lime Survey* are currently being used, along with a range of other tools that are hosted both on- and off-site. To determine whether or not this was common practice, we examined each of the studies employing online survey methods listed on Hanover College’s *Psychological Research on the Net* website on 19 September 2009 that had been added in the three months from 20 June to 19 September 2009. *Psychological Research on the Net* was selected because of its size, popularity, and exclusive focus on ethical academic research (the requirements for listing a study on the site include providing information about the researchers, affiliations, and ethics review processes).

Of the 66 studies meeting our criteria, 35 had chief investigators (CIs) with affiliations at United States universities or colleges, and 23 had CIs with United Kingdom affiliations. The remaining studies were Australian (4), Canadian (1), Irish (1), Singaporean (1) and Swiss (1).

Consistent with Buchanan and Hvizdak (2009), who found that just 24% of the United States Human Research Ethics Committee (HREC) representatives they surveyed worked at institutions with “specific tool[s] to use for online surveys” (p. 40), only 17 (i.e., 26%) of the 66 surveys we examined were hosted on web-servers owned and operated by the CI’s institution, or another academic institution with which the CI was affiliated. Of the remaining 49 surveys, 47 were hosted off-site (see Table 1), and we were unable to draw any conclusions about the final two. Excluding the five surveys hosted on personally owned web servers, the

off-site surveys we looked at were exclusively hosted by commercial service providers, primarily *SurveyMonkey*.

These findings suggest considerable variation across institutions and researchers, with the majority outsourcing major aspects of the web surveying process to commercial service providers. Such outsourcing can offer a number of advantages to academic researchers. First, it is typically quicker and easier to use existing products for survey design and deployment, than to develop systems internally. Ease of use may be of particular concern to academics supervising student research projects with short time-lines, or utilising online surveys in their teaching (Gaiser & Schreiner, 2009). Second, outsourcing usually eliminates the need for sophisticated technical knowledge, including the need to maintain a web-server and databases (Kaczmirek, 2008). Furthermore, large commercial providers can usually offer researchers guaranteed ‘up-time’, a regular backup schedule, and high levels of data security (Kaczmirek, 2008), often at a considerably lower cost than deploying and maintaining a comparable service in-house (Gaiser & Schreiner, 2009; Kaczmirek, 2008). On the surface, these advantages make the outsourcing of web surveys an attractive option for many researchers. However, outsourcing also raises a number of significant ethical concerns.

ETHICAL ISSUES ASSOCIATED WITH OUTSOURCING

In the previous section, we noted the popularity of outsourcing significant aspects of the web surveying process to external (and typically commercial) service providers. Such outsourcing can offer many advantages, but also raises a number of ethical concerns, particularly when service providers are selected and used by researchers on a seemingly case-by-case, ad-hoc basis¹⁰. In this section we examine ethical issues associated with outsourcing, focusing on two key areas. First, we outline potential threats to anonymity and confidentiality associated with

Table 1. Hosting locations of 47 online surveys listed at 'Psychological Research on the Net' in the three months to 19 September 2009 and not hosted on the CI's Institution's web-servers

Host	Website address	N
SurveyMonkey	http://surveymonkey.com	27
Psych Data	https://psychdata.com	5
Qualtrics	http://qualtrics.com	2
Survey Gizmo	http://surveygizmo.com	2
Bristol Online Surveys	http://survey.bris.ac.uk	2
Globalpark/Unipark	http://unipark.info	2
Formsite	http://formsite.com	1
QuestionPro	http://questionpro.com	1
Researcher's Personal Web Server	n/a	5
Total		47

both data protection methods and the collection of IP addresses. Then we examine the potential impact of the perceived credibility of a data collection website on response rates and the accuracy of reporting. While recognising that each discipline has their own set of ethical guidelines, in our discussion of these issues we refer to the American Psychological Association's Ethical Guidelines (APA, 2002). These guidelines, in common with most other sets of ethical guidelines, are based on the principles of beneficence and nonmaleficence, fidelity and responsibility, integrity, justice and respect for the rights and dignity of individuals.

Data Protection: Threats to Anonymity and Confidentiality

The protection of data at all stages of the research process, from initial data collection through to storage, is vital to ensuring the confidentiality and anonymity of research participants. With online research, data protection moves beyond the traditional methods for protection of paper documents to cover the protection of digital data. The potential for intentional malicious damage to online surveys is not simply a theo-

retical risk. Online surveys have been hacked (see Andrews, Nonnecke, & Preece, 2003, for details of how their online survey was hacked twice and infected with a virus) highlighting the need to ensure a high level of data protection. As noted by the American Psychological Association Policy and Planning Board (2009) "issues of protecting participant privacy in Internet transmission and computer storage are paramount but challenging" (p. 458).

The data protection measures employed need to increase with the increasing sensitivity of the data collected. Barchard and Williams (2008) recommended researchers of highly sensitive topics go beyond basic security measures and refer to the security standards in the computing industry, such as those provided by the Payment Card Industry Standards Council¹¹, for the most up-to-date advice on data protection. The American Psychological Association's Board of Scientific Affairs' Advisory Group go further, recommending that where acceptable protections cannot be put in place, alternatives to Internet research should be used (Kraut, Olson, Banaji, Bruckman, Cohen, & Couper, 2004).

The outsourced hosting of surveys is associated with additional layers of threats to

data protection over those shared by all web surveys. While many commercial web survey hosting services may employ high level data protection measures that are consistent with industry standards¹², a major concern is that the researcher does not have complete control over who can, and cannot, access the research data. A second area of concern with the external hosting of web surveys is the additional risks associated with the transmission of data from the host to the researcher.

External hosting services also vary in their data protection policies and practices. Further highlighting the potential for breaches of data security, Buchanan and Hvizdak (2009) reported that more than a third of their Human Research Ethics Committee representative survey respondents did not, as part of the ethics review process, consider the security and privacy policies of external service providers. As Buchanan and Hvizdak noted (2009), “until each tool is vetted and its privacy policies and data security policies understood, we cannot be 100% certain how security, content and privacy are instantiated within the individual tools” (p. 46).

Collection of IP Addresses: A Threat to Anonymity

A further threat to participant anonymity is the collection of IP addresses. A unique Internet Protocol (IP) address is assigned to a computer each time it connects to the Internet. Banks of IP addresses are allocated to organisations and Internet Service Providers (ISPs) through five regional Internet registries: AfriNIC servicing the Africa region, APNIC (Asia Pacific), LACNIC (Latin America and the Caribbean), American Registry for Internet Numbers (ARIN) and RIPE NCC covering Europe, the Middle East and parts of Central Asia. In some circumstances it is possible to trace the location of a specific computer from an IP address. This may be done through one of the regional registries, along with the records of the ISP originally allocated the address of interest (Barchard & Williams, 2008).

While it is possible to use IP addresses and cookies to identify/track use on individual computers (Charlesworth, 2008), it is difficult to make a definitive link from an IP address to a specific individual. An IP address only identifies a computer, not a user (Nosek, Banaji, & Greenwald, 2002). Furthermore, many ISPs use dynamic IP allocation, whereby an IP address is assigned to a computer for the duration of the session only (Nosek et al., 2002), meaning that over a course of a day several computers may have been assigned the same IP address. Furthermore, a computer may be used by multiple users (e.g., a computer located in a public library) and/or a single account may be used by multiple family members (Hewson, Yule, Laurent, & Vogel, 2003).

However, the uniqueness of IP addresses, when used in combination with time and date information, means they should be treated in survey research as potential identifiers. Preferably, IP addresses should not be recorded as part of a survey (Nosek et al., 2002). When using an external survey provider, the option of not recording IP addresses may not be possible. Where a commercial survey provider automatically captures IP addresses, it is recommended that they be deleted as soon as possible, preferably before saving the data file to the researcher’s computer (Barchard & Williams, 2008; Benfield & Szlemko, 2006). However, the external survey provider is likely to retain IP information, regardless of whether or not the researcher deletes it, posing an ongoing threat to confidentiality and anonymity. For example, the *SurveyMonkey* Privacy Policy¹³ states:

As is true of most Web sites, we gather certain information automatically and store it in log files. This information includes internet protocol (IP) addresses, browser type, internet service provider (ISP), referring/exit pages, operating system, date/time stamp, and clickstream data.

We use this information, which does not identify individual users, to analyze trends, to administer the site, to track users’ movements around the

site and to gather demographic information about our user base as a whole.

We do not link this automatically-collected data to personally identifiable information.

However, that such data is not generally linked does not mean it will never be linked. Later in the *SurveyMonkey* Privacy Policy under 'Legal Disclosure' it is stated that:

We reserve the right to disclose your personally identifiable information as required by law and when we believe that disclosure is necessary to protect our rights and/or to comply with a judicial proceeding, court order, or legal process served on our Web site.

This effectively means that control over the decision of whether or not to disclose research data to legal authorities may be taken out of the hands of the researcher and his/her institution. This may be a particular issue for researchers conducting surveys on criminal behaviour, where there have been cases of off-line research data being subpoenaed or research suspended over concerns about being able to maintain confidentiality (Roberts & Indermaur, 2003). In line with the APA's recommendations on informed consent (APA, 2002), research participants must be informed of the limits of confidentiality.

The Impact of Credibility of Site on Response Rates and Accuracy of Reporting

Ethical issues also arise in relation to public perceptions of the credibility of surveys hosted at non-academic domains. The external hosting of an academic web survey risks diluting public perceptions of the academic nature of the research. In addition to academic researchers, commercial, non-profit and media organisations, and members of the lay-public also use web surveys to collect data. For example, Couper (2000) refers to 'web surveys as entertainment', which includes collections

of non-scientific surveys or polls and media 'question of the day' polls. Some potential research participants may be unable to differentiate between academic research surveys and other commercial surveys, potentially affecting the credibility of academic surveys housed by commercial survey providers (Binik, Mah, & Kiesler, 1999; Fricker & Schonlau, 2002). Some external hosting services routinely use banner advertisements on survey pages, further blurring the distinction between academic and commercial data collection. This highlights the need for researchers to clearly delineate their work as 'academic research' that has ethical approval from the relevant HRECs/IRBs.

Suggested ways of strengthening the perceived links between research and academic institutions include posting researchers' photographs and links to researchers' home pages on the survey site (Binik et al., 1999). Peden and Flashinski (2004) examined psychology research websites for evidence of institutional affiliation. Only 22% of 22 websites housing psychology surveys and experiments reviewed in early 2002 contained an active link to a university website, although 88% identified institutional affiliations. Further, only a minority of sites (31%) stated that the research had been granted ethical approval by a HREC/IRB, with even fewer (27%) actually providing contact details for the approving body.

The perceived credibility of a survey domain may affect both willingness to participate in research and the candidacy of responding. While Internet users vary in their levels of concern about online privacy, the majority do express some concern about disclosing personal information online. For example, of 1,482 US residents surveyed as part of an online survey about Internet use, 53.7% reported being 'very concerned' and 27.1% 'somewhat concerned' about security on the Internet, where security was defined to include privacy, confidentiality and identity issues (O'Neil, 2001). Further, online privacy concern may vary by domain. Home Internet users vary in the degree to which they find website privacy statements from corporations and government institutions credible

(Turow & Hennessy, 2007). While the proportion of Internet users who trust commercial online survey providers or universities has not been established, the percentage of 1,200 adult home Internet users surveyed who trusted an institution to protect their information online and not disclose it without their consent varied by institutional type, from 4% for major advertisers to 25% for makers of privacy protection software (Turow & Hennessy, 2007).

The presence of online privacy policies on websites has limited impact on perceptions of privacy risk (Myerscough, Lowe, & Alpert, 2006). Further, the majority of Internet users do not systematically read online privacy notices. Based on survey responses from a stratified random sample of 2,468 U.S. adults from the Harris Poll Online panel, Milne and Culnan (2004) reported that 17.3% of respondents stated they never read privacy notices on websites. Of those who did report reading privacy notices, less than five percent reported always reading them. As Binik et al. (1999) suggest, "researchers should not assume that a promise of anonymity or non-anonymity is always viewed as such by participants" (pp. 85-86).

Where individuals have online privacy concerns, the majority take actions to protect their privacy (Paine, Reips, Steiger, Joinson, & Buchanan, 2007). While protective measures are largely based around hardware and software (e.g., firewalls, use of antivirus software etc.), almost 10% of Paine and colleagues' survey respondents volunteered that they were careful about the information they revealed online. Experimental research suggests that online survey responding is sensitive to, and responses may be affected by, privacy concerns. Joinson, Paine, Buchanan, and Reips (2008) manipulated level of privacy concern in online surveys, demonstrating that the use of an 'I prefer not to say' option is sensitive to both priming and manipulation of privacy concern.

Online privacy concern may also affect the candidness of survey responses. While early research into the computer administration of measures suggested that this mode of administration reduced socially desirable responding

and increased the candidness of responses (Feigelson & Dwight, 2000), more recent research has failed to find differences between various modes of administration (e.g., Bates & Cox, 2008; Uriell & Dudley, 2009). Respondent concerns over web survey data security have the potential to reverse any positive effects on social desirability responding (Couper, 2000).

Perceptions of confidentiality and anonymity of survey responses can affect responding to survey questions deemed sensitive by the respondent. A meta-analysis of research conducted into the effect of confidentiality assurances in offline research indicated that confidentiality assurances can improve responding to sensitive questions (Singer, 2004; Singer, Von Thurn, & Miller, 1995). More recent research has suggested that perceptions of anonymity have a greater effect than assurances of confidentiality on preparedness to reveal sensitive information (Ong & Weiss, 2000).

In addition to the impact of the immediate environment, Binik et al. (1999) suggest that online cues and the survey interface may impact on perceptions of anonymity. Perceptions of anonymity and security of survey responses influence intention to respond to online surveys (Rogelberg, Spitzmueller, Little, & Reeve, 2006) and accuracy of reporting. Uriell and Dudley's (2009) survey of enlisted US navy personnel found that web survey respondents were significantly more likely than pen-and-paper survey respondents to think that others could access their survey responses and that their survey responses would be linked with identifying and personal information. Accuracy of responses was positively correlated with perceived anonymity and confidentiality of survey responses. Participants' concern over the potential identifiability of data from web surveys suggests that researchers need to make explicit how anonymity will be maintained (Chizawsky, Estabrooks, & Sales, 2009).

The history of privacy violations online creates an atmosphere un conducive to building a relationship of trust between respondents and researchers (Cho & LaRose, 1999). This distrust may be magnified where commercial

survey providers are utilised for data collection. Research in offline settings has demonstrated that the perceived legitimacy and authority of researchers is influential in the decision to participate in research (Groves, Cialdini, & Couper, 1992) with higher responses rates for university sponsored research (Fox, Crask, & Kim, 1998). The internal hosting of web surveys on education domains may increase the credibility of research and hence response rates (Cho & LaRose, 1999), as well as the candidness of responding.

HOSTING ON-SITE

Researchers may seek to avoid or address some of the ethical concerns associated with outsourcing by simply moving their web surveying on-site. This can seem particularly tempting to those researchers with a reasonable degree of IT savvy and administrator level access to a web server. We do not wish to imply that the outsourcing of academic web survey development and hosting is necessarily inferior to developing and hosting surveys internally. Indeed, while internal development and hosting increases the transparency of research (Buchanan & Hvizdak, 2009) and strengthens the identification of the research with the university, it can also raise a raft of new concerns. For example, are procedures in place to ensure that the both the surveying application *and* the software and services on which it relies (e.g., the web server, database server, web application framework etc.) are appropriately maintained (i.e., regularly updated/patched, backed-up etc.)? How are 'default' security and privacy policies set, and reviewed? Who has administrator level access to the web server, and are these people appropriately qualified? How are access rights and user accounts managed? Can users edit and/or view each other's surveys or data? If so, how is confidentiality managed? These issues are largely beyond the scope of this paper, but illustrate that the decision about whether to outsource or not is a challenging one, and should not be made lightly. With this

in mind, in the final section of this paper, we offer a series of suggestions to those readers needing to make such a decision.

GUIDELINES

First and foremost, we recommend that each university develop a coordinated, institution-wide approach to online surveying, rather than relying on ad-hoc decisions by individual researchers, and the duplication of systems and services that such decisions often result in. We recommend the development of this approach involve representatives from the university HREC/IRB, legal department and IT department, in addition to academics from a range of disciplines who are experienced in conducting online research. A set of clearly stated policies and procedures for conducting web surveying should also be developed. As part of a coordinated, institution-wide approach, a university may choose to provide and support internal survey development and hosting and/or to provide a short-list of 'approved' external services for survey development and hosting. Each of these options will be briefly explored below.

In our view, the greatest protection to research participants is offered where the university provides and supports the development and hosting of online surveys, and the online surveying facilities are managed and maintained by staff skilled in IT security and familiar with the ethical and legal requirements that researchers are bound by in their geographic regions and professional disciplines. Such facilities can be based on an open source software package like *Lime Survey*, or a proprietary solution such as *Opinio*. Larger institutions may also consider the option of developing a customised surveying package in-house, rather than depending on code developed or maintained by outsiders.

However, we recognise that it is not always possible to harness the resources necessary to provide surveying facilities in-house. This may be particularly the case for smaller or specialised institutions, or institutions where there is little demand for web surveying. Where this is the

case, we would recommend that representatives from the university HREC, legal department, IT department and active research academics examine the terms of use and security provisions of a range of widely used commercial survey providers with the aim of providing a short list of acceptable providers. In recognition of the rapidly changing field, it is recommended that this list of preferred providers be reviewed on an annual basis. Where necessary for the specifics of their research project, individual students/researchers can present a case for utilising another survey organisation, and this can be assessed on a case-by-case basis.

Where the decision is made to outsource the hosting of a survey, we recommend that the survey content, hosted on the commercial site, is 'sandwiched' between an information sheet and debriefing page, both hosted on a university server. This will strengthen perceptions of the association between the research and the university. It also allows for the collection of identifying information for purposes such as informed consent or entry into a prize draw to occur on the university server. This separation of collecting survey information on the commercial survey provider's server and identifying information on a university server provides an additional layer of protection for participants (Barchard & Williams, 2008).

Where a university has not developed a coordinated, institution-wide approach to online surveying, individual researchers may need to make their own decisions about outsourcing aspects of their web survey research. In our own research and supervision of research students we have successfully used both internally hosted surveys developed using an open source software package and surveys externally hosted on commercial web surveying sites. These choices were largely influenced by the technical skills and experience of the researchers/students and duration of the projects, with those with limited IT skills and a limited data collection period being directed towards external survey companies where the researcher requires few technical skills to be able to 'create' their

on-line survey. In choosing between external providers, particular consideration should be given to data protection and privacy policies, privacy certification, and hardware and software configurations.

CONCLUSION

The use of web surveying in academic research is a relatively new phenomenon, and occurs within a rapidly changing environment characterized by technological innovation. New modes of data collection are likely to evolve, enabled by technological change (Tourangeau, 2004). While the principles underlying ethical research remain the same, the application of these principles to new methodologies such as web surveying lags behind their introduction. In this article we have outlined some of the ethical issues associated with outsourcing aspects of web surveying at the current point in time. While we have provided suggested guidelines in relation to the outsourcing (or otherwise) of web surveys, researchers will need to keep abreast of both social and technological changes in the field, including both standards for data protection and evolving interpretations of ethical codes.

REFERENCES

- American Psychological Association. (2002). *Ethical principles of psychologists and code of conduct*. Retrieved from <http://www.apa.org/ethics/code2002.pdf>
- American Psychological Association Policy and Planning Board. (2009). How technology changes everything (and nothing) in psychology: 2008 annual report of the APA Policy and Planning Board. *American Psychologist*, 64, 454-463. doi:10.1037/a0015888
- Andrews, D., Nonnecke, B., & Preece, J. (2003). Electronic survey methodology: A case study in reaching hard-to-involve Internet users. *International Journal of Human-Computer Interaction*, 16, 185-210. doi:10.1207/S15327590IJHC1602_04

- Australian Bureau of Statistics. (2008). *Household use of information technology, Australia, 2007-08* (Catalogue Number 8146.0). Canberra, Australia: Author. Retrieved from <http://www.abs.gov.au/Ausstats/abs@.nsf/mf/8146.0>
- Baker, J. D. (2007). Open source survey software . In Reynolds, R. A., Woods, R., & Baker, J. D. (Eds.), *Handbook of research on electronic surveys and measurements* (pp. 273–275). Hershey, PA: IGI Global.
- Barchard, K. A., & Williams, J. (2008). Practical advice for conducting ethical online experiments and questionnaires for United States psychologists. *Behavior Research Methods, 40*, 1111–1128. doi:10.3758/BRM.40.4.1111
- Bates, S. C., & Cox, J. M. (2008). The impact of computer versus paper-pencil survey, and individual versus group administration, on self-reports of sensitive behaviors. *Computers in Human Behavior, 24*, 903–916. doi:10.1016/j.chb.2007.02.021
- Beiderniki, G., & Kerschbaumer, A. (2007). Comparison of online survey tools. In Reynolds, R. A., Woods, R., & Baker, J. D. (Eds.), *Handbook of research on electronic surveys and measurements* (pp. 264–272). Hershey, PA: IGI Global.
- Benfield, J. A., & Szlemko, W. J. (2006). Internet-based data collection: Promises and realities. *Journal of Research Practice, 2*, 1–15.
- Best, S. J., & Krueger, B. S. (2004). *Internet data collection*. Thousand Oaks, CA: Sage.
- Best, S. J., & Krueger, B. S. (2008). Internet survey design. In Fielding, N., Lee, R. M., & Blank, G. (Eds.), *The SAGE handbook of online research methods* (pp. 217–235). London: Sage.
- Binik, Y. M., Mah, K., & Kiesler, S. (1999). Ethical issues in conducting sex research on the Internet. *Journal of Sex Research, 36*, 82–90. doi:10.1080/00224499909551971
- Birnbaum, M. H. (2007). Designing online experiments . In Joinson, A., McKenna, K., Postmes, T., & Reips, U.-D. (Eds.), *The Oxford handbook of Internet psychology* (pp. 391–403). New York: Oxford University Press.
- Buchanan, E. A., & Hvizdak, E. E. (2009). Online survey tools: Ethical and methodological concerns of Human Research Ethics Committees. *Journal of Empirical Research on Human Research Ethics; JERHRE, 4*, 37–48. doi:10.1525/jer.2009.4.2.37
- Charlesworth, A. (2008). Understanding and managing legal issues in Internet research. In Fielding, N., Lee, R. M., & Blank, G. (Eds.), *The SAGE handbook of online research methods* (pp. 42–57). London: Sage.
- Chizawsky, L. L. K., Estabrooks, C. A., & Sales, A. E. (in press). The feasibility of web-based surveys as a data collection tool: A process evaluation. *Applied Nursing Research*.
- Cho, H., & LaRose, R. (1999). Privacy issues in Internet surveys. *Social Science Computer Review, 17*, 421–434. doi:10.1177/089443939901700402
- Couper, M. P. (2000). Web surveys: A review of issues and approaches. *Public Opinion Quarterly, 64*, 464–494. doi:10.1086/318641
- Crawford, S. (2002). Evaluation of web survey data collection systems. *Field Methods, 14*, 307–321.
- Directorate for Science, Technology and Industry, Organisation for Economic Co-Operation and Development. (2009). *OECD key ICT indicators*. Retrieved September 20, 2009, from <http://www.oecd.org/sti/ICTindicators>
- Ess, C. (2007). Internet research ethics . In Joinson, A., McKenna, K., Postmes, T., & Reips, U.-D. (Eds.), *The Oxford handbook of Internet psychology* (pp. 487–502). New York: Oxford University Press.
- Ess, C., & the The AIOR Ethics Working Committee. (2002). *Ethical decision making and Internet research: Recommendations from the AOIR ethics working committee*. Retrieved August 31, 2009, from <http://www.aoir.org/reports/ethics.pdf>
- Evans, J. R., & Mathur, A. (2005). The value of online surveys. *Internet Research, 15*, 195–219. doi:10.1108/10662240510590360
- Feigelson, M. E., & Dwight, S. A. (2000). Can asking questions by computer improve the candidness of responding? A meta-analytic perspective. *Consulting Psychology Journal: Practice and Research, 52*, 248–255. doi:10.1037/1061-4087.52.4.248
- Fox, R. J., Crask, M. R., & Kim, J. (1988). Mail survey response rate: A meta-analysis of selected techniques for inducing response. *Public Opinion Quarterly, 52*, 467–491. doi:10.1086/269125
- Fricke, R. D. Jr, & Schonlau, M. (2002). Advantages and disadvantages of Internet research surveys: Evidence from the literature. *Field Methods, 14*, 347–367. doi:10.1177/152582202237725

- Gaiser, T. J., & Schreiner, A. E. (2009). *A guide to conducting online research*. London: Sage.
- Gordon, A. (2002). Product review: SurveyMonkey.com—Web-based survey and evaluation system. *The Internet and Higher Education*, 5, 83–87. Retrieved from <http://www.SurveyMonkey.com>. doi:10.1016/S1096-7516(02)00061-1
- Gosling, S. D., Vazire, S., Srivastava, S., & John, O. P. (2004). Should we trust web-based studies? A comparative analysis of six preconceptions about Internet questionnaires. *The American Psychologist*, 59, 93–104. doi:10.1037/0003-066X.59.2.93
- Greenfield, D. N. (1999). Psychological characteristics of compulsive Internet use: A preliminary analysis. *Cyberpsychology & Behavior*, 2, 403–412. doi:10.1089/cpb.1999.2.403
- Groves, R. M., Cialdini, R. B., & Couper, M. P. (1992). Understanding the decision to participate in a survey. *Public Opinion Quarterly*, 56, 475–495. doi:10.1086/269338
- Hewson, C. (2007). Gathering data on the Internet: Qualitative approaches and possibilities for mixed methods research. In Joinson, A., McKenna, K., Postmes, T., & Reips, U.-D. (Eds.), *The Oxford handbook of Internet psychology* (pp. 405–428). New York: Oxford University Press.
- Hewson, C., & Laurent, D. (2008). Research design and tools for Internet research. In Fielding, N., Lee, R. M., & Blank, G. (Eds.), *The SAGE handbook of online research methods* (pp. 58–78). London: Sage.
- Hewson, C., Yule, P., Laurent, D., & Vogel, C. (2003). *Internet research methods: A practical guide for the social and behavioural sciences*. London: Sage.
- Hildebrandt, T., Langenbacher, J., Carr, S., Sanjuan, P., & Park, S. (2006). Predicting intentions for long-term anabolic-androgenic steroid use among men: A covariance structure model. *Psychology of Addictive Behaviors*, 20, 234–240. doi:10.1037/0893-164X.20.3.234
- Horrigan, J. (2009, June). *Home broadband adoption 2009: Broadband adoption increases, but monthly prices do too*. Washington, DC: Pew Internet & American Life Project. Retrieved from <http://pewinternet.org/Reports/2009/10-Home-Broadband-Adoption-2009.aspx>
- Jenetzko, D. (2008). Nonreactive data collection on the Internet. In Fielding, N., Lee, R. M., & Blank, G. (Eds.), *The SAGE handbook of online research methods* (pp. 161–173). London: Sage.
- John, O. P., Naumann, L. P., & Soto, C. J. (2008). Paradigm shift to the integrative big-five trait taxonomy: History, measurement, and conceptual Issues. In John, O. P., Robins, R. W., & Pervin, L. A. (Eds.), *Handbook of personality: Theory and research* (3rd ed., pp. 114–158). New York: Guilford Press.
- Joinson, A. N., Paine, C., Buchanan, T., & Reips, U.-D. (2008). Measuring self-disclosure online: Blurring and non-response to sensitive items in web-based surveys. *Computers in Human Behavior*, 24, 2158–2171. doi:10.1016/j.chb.2007.10.005
- Kaczmirek, L. (2008). Internet survey software tools. In Fielding, N., Lee, R. M., & Blank, G. (Eds.), *The SAGE handbook of online research methods* (pp. 236–254). London: Sage.
- Keller, H. E., & Lee, S. (2003). Ethical issues surrounding human participants research using the Internet. *Ethics & Behavior*, 13, 211–219. doi:10.1207/S15327019EB1303_01
- Kraut, R., Olson, J., Banaji, M., Bruckman, A., Cohen, J., & Couper, M. (2004). Psychological research online: Report of the Board of Scientific Affairs' Advisory Group on the conduct of research on the Internet. *The American Psychologist*, 59, 105–117. doi:10.1037/0003-066X.59.2.105
- Lee, R. M., Fielding, N., & Blank, G. (2008). The Internet as a research medium: An editorial introduction to 'The Sage Handbook of Online Research Methods'. In Fielding, N., Lee, R. M., & Blank, G. (Eds.), *The SAGE handbook of online research methods* (pp. 3–20). London: Sage.
- Luppardini, R. (2009). The emerging field for technoethics. In Luppardini, R., & Adell, R. (Eds.), *Handbook of research on technoethics* (pp. 1–19). Hershey, PA: IGI Global.
- McFarlane, M., Bull, S. S., & Rietmeijer, C. A. (2002). Young adults on the Internet: Risk behaviors for sexually transmitted diseases and HIV. *The Journal of Adolescent Health*, 31, 11–16. doi:10.1016/S1054-139X(02)00373-7
- Milne, G. R., & Culnan, M. J. (2004). Strategies for reducing online privacy risks: Why consumers read (or don't read) online privacy notices. *Journal of Interactive Marketing*, 18(3), 15–29. doi:10.1002/dir.20009
- Myerscough, S., Lowe, B., & Alpert, F. (2006). Willingness to provide personal information online: The role of perceived privacy risk, privacy statements and brand strength. *Journal of Website Promotion*, 2, 115–139. doi:10.1080/15533610802104182

- Nosek, B. A., Banaji, M. R., & Greenwald, A. G. (2002). E-research: Ethics, security, design and control in psychological research on the Internet. *The Journal of Social Issues, 58*, 161–176. doi:10.1111/1540-4560.00254
- O'Connor, H., Madge, C., Shaw, R., & Wellens, J. (2008). Internet-based interviewing. In Fielding, N., Lee, R. M., & Blank, G. (Eds.), *The SAGE handbook of online research methods* (pp. 271–289). London: Sage.
- O'Neil, D. (2001). Analysis of Internet users' level of online privacy concern. *Social Science Computer Review, 19*, 17–31. doi:10.1177/089443930101900103
- Office for National Statistics. (2009, August 28). *Internet access: Households and individuals*. Retrieved from <http://www.statistics.gov.uk/pdfdir/iah0809.pdf>
- Ong, A. D., & Weiss, D. J. (2000). The impact of anonymity on responses to sensitive questions. *Journal of Applied Social Psychology, 30*, 1691–1708. doi:10.1111/j.1559-1816.2000.tb02462.x
- Paine, C., Reips, U.-D., Stieger, S., Joinson, A., & Buchanan, T. (2007). Internet users' perceptions of 'privacy concerns' and 'privacy actions'. *International Journal of Human-Computer Studies, 65*, 526–536. doi:10.1016/j.ijhcs.2006.12.001
- Peden, B. F., & Flashinski, D. P. (2004). Virtual research ethics: A content analysis of surveys and experiments online. In Buchanan, E. (Ed.), *Readings in virtual research ethics: Issues and controversies* (pp. 1–26). Hershey, PA: IGI Global.
- Reips, U.-D. (2007). The methodology of Internet-based experiments. In Joinson, A., McKenna, K., Postmes, T., & Reips, U.-D. (Eds.), *The Oxford handbook of Internet psychology* (pp. 373–390). New York: Oxford University Press.
- Reips, U.-D. (2008). How Internet-mediated research changes science. In A. Barak (Ed.), *Psychological aspects of cyberspace: Theory, research, applications* (pp. 268–294). Cambridge, UK: Cambridge University Press. Retrieved from <http://gsb.haifa.ac.il/~sheizaf/cyberpsych/12-Reips.pdf>
- Rentfrow, P. J., Gosling, S. D., & Potter, J. (2008). A theory of the emergence, persistence, and expression of geographic variation in psychological characteristics. *Perspectives on Psychological Science, 3*, 339–369. doi:10.1111/j.1745-6924.2008.00084.x
- Roberts, L., & Indermaur, D. (2003). Signed consent forms in criminological research: Protection for researchers and ethics committees but a threat to research participants? *Psychiatry, Psychology and Law, 10*, 289–299. doi:10.1375/pplt.2003.10.2.289
- Robins, R. W., Tracy, J. L., Trzesniewski, K., Gosling, S. D., & Potter, J. (2002). Global self-esteem across the life span. *Psychology and Aging, 17*, 423–434. doi:10.1037/0882-7974.17.3.423
- Robins, R. W., Tracy, J. L., Trzesniewski, K., Potter, J., & Gosling, S. D. (2001). Personality correlates of self-esteem. *Journal of Research in Personality, 35*, 463–482. doi:10.1006/jrpe.2001.2324
- Robinson, K. M. (2001). Unsolicited narratives from the Internet: A rich source of qualitative data. *Qualitative Health Research, 11*, 706–714. doi:10.1177/104973201129119398
- Rogelberg, S. C., Spitzmueller, C., Little, I., & Reeve, C. L. (2006). Understanding response behavior to an online special topics organizational satisfaction survey. *Personnel Psychology, 59*, 903–923. doi:10.1111/j.1744-6570.2006.00058.x
- Singer, E. (2004). Confidentiality, risk perception and survey participation. *Chance, 17*(3), 31–35.
- Singer, E., Von Thurn, D. R., & Miller, E. R. (1995). Confidentiality assurances and response: A quantitative review of the experimental literature. *Public Opinion Quarterly, 59*, 66–77. doi:10.1086/269458
- Skitka, L. J., & Sargis, E. G. (2006). The Internet as psychological laboratory. *Annual Review of Psychology, 57*, 529–555. doi:10.1146/annurev.psych.57.102904.190048
- Soto, C. J., John, O. P., Gosling, S. D., & Potter, J. (2008). The developmental psychometrics of big five self-reports: Acquiescence, factor structure, coherence, and differentiation from ages 10 to 20. *Journal of Personality and Social Psychology, 94*, 718–737. doi:10.1037/0022-3514.94.4.718
- Srivastava, S., John, O. P., Gosling, S. D., & Potter, J. (2003). Development of personality in early and middle adulthood: Set like plaster or persistent change? *Journal of Personality and Social Psychology, 84*, 1041–1053. doi:10.1037/0022-3514.84.5.1041
- Sue, V. M., & Ritter, L. A. (2007). *Conducting online surveys*. Thousand Oaks, CA: Sage.
- Tourangeau, R. (2004). Survey research and societal change. *Annual Review of Psychology, 55*, 775–801. doi:10.1146/annurev.psych.55.090902.142040

Turow, J., & Hennessy, M. (2007). Internet privacy and institutional trust: Insights from a national survey. *New Media & Society*, 9, 300–318. doi:10.1177/1461444807072219

Uriell, Z. A., & Dudley, C. M. (2009). Sensitive topics: Are there modal differences? *Computers in Human Behavior*, 25, 76–87. doi:10.1016/j.chb.2008.06.007

Whang, L. S.-M., Lee, S., & Chang, G. (2003). Internet over-users' psychological profiles: A behavior sampling analysis on Internet addiction. *Cyberpsychology & Behavior*, 6, 143–150. doi:10.1089/109493103321640338

Wright, K. B. (2005). Researching Internet-based populations: Advantages and disadvantages of online survey research, online questionnaire authoring software packages, and web survey services. *Journal of Computer-Mediated Communication*, 10(3). Retrieved from <http://jcmc.indiana.edu/vol10/issue3/wright.html>.

ENDNOTES

- 1 A recent *Google Scholar* search by Lee et al. (2008) indicated that the number of social science articles with 'web survey', 'Internet survey' or 'online survey' in their titles increased from 4 in 1994 to 1146 in 2006.
- 2 <http://psych.hanover.edu/Research/exponnet.html>
- 3 <http://genpsylab-wexlist.unizh.ch/browse.cfm?action=browse&modus=survey>
- 4 <http://surveymonkey.com/>

WebSM (<http://websm.org/>) allows users to search through 350+ web surveying applications and services on characteristics like cost to the user, availability of source code (i.e., closed vs. open source), and whether or not the user's surveys and data are hosted on the vendor's, or user's own web-server.

<http://www.adobe.com/products/dreamweaver/>
<http://www.microsoft.com/expression/>
<http://www.objectplanet.com/opinio/>; a proprietary application developed and distributed by Object Planet Inc.
<http://www.limesurvey.org/>

This paper focuses solely on ethical issues associated with outsourcing web surveys. That is, the use of commercial survey hosting services for academic surveys. For a more general discussion of online research ethics please see Ess (2007) and Ess and the AIOR Ethics Working Committee (2002). Our focus on the ethical issues associated with a specific online methodology and context is consistent with Ess's (2007) claim that "research ethics is intimately interwoven with the specific methodology/ies used in a given project" (p. 495).

- 11 See https://www.pcisecuritystandards.org/security_standards/pci_dss.shtml
- 12 For example, *SurveyMonkey* is a licensee of the TRUSTe Privacy Program, complies with the EU Safe Harbor framework and employs Secure Socket Layer (SSL) technology to encrypt sensitive information.
- 13 http://www.surveymonkey.com/Monkey_Privacy.aspx (last accessed on 21 September 2009).

Peter J. Allen is a PhD candidate and Associate Lecturer in the School of Psychology and Speech Pathology at Curtin University in Perth, Western Australia and member of the Curtin Health Innovation Research Institute. He is interested in the psychology of Internet behaviour, and is currently investigating the psychological aspects of online copyright infringement. Peter is co-author of the popular SPSS text, PASW Statistics by SPSS: A Practical Guide.

Lynne D. Roberts is a Senior Lecturer in the School of Psychology and Speech Pathology at Curtin University in Perth, Western Australia and member of the Curtin Health Innovation Research Institute. Lynne has conducted and published research within the field of technoethics, with a particular focus on cyber-crime, cyber-victimisation and online research ethics. Her other research interests centre on social interaction online, online research methodologies and public attitudes to crime and justice. Details of Lynne's publications in these areas are available at http://psych.curtin.edu.au/staff_profile.cfm?id=25

**Paper 7: Exploring Ethical Issues Associated with Using Online Surveys in
Educational Research**

Roberts, L. D., & Allen, P. J. (2015). Exploring ethical issues associated with using online surveys in educational research. *Educational Research and Evaluation*, 21, 95-108. doi:10.1080/13803611.2015.1024421

This is an Accepted Manuscript of an article published by Taylor & Francis in Educational Research and Evaluation on 14 Apr 2015, available online: <http://www.tandfonline.com/10.1080/13803611.2015.1024421>

In compliance with Taylor & Francis's terms and conditions of use, this is the post-print version of the abovementioned work. Readers wishing to cite this paper are encouraged to source the final published version, available at the URL above.

Exploring ethical issues associated with using online surveys in educational research

Lynne D. Roberts and Peter J. Allen

School of Psychology and Speech Pathology, Curtin University, Australia

Corresponding Author:

Associate Professor Lynne Roberts

School of Psychology and Speech Pathology

Curtin University

GPO Box U1987

Perth WA 6845

Telephone: +61 8 9266 7183

Lynne.Roberts@curtin.edu.au

Peter Allen

School of Psychology and Speech Pathology

Curtin University

GPO Box U1987

Perth WA 6845

Telephone: +61 8 9266 2554

P.Allen@curtin.edu.au

Disclosure Statement:

The authors of this publication received no financial interest or benefit from this research.

Abstract

Online surveys are increasingly used in educational research, yet little attention has focused on ethical issues associated with their use in educational settings. Here, we draw on the broader literature to discuss five key ethical issues in the context of educational survey research: dual teacher/researcher roles; informed consent; use of incentives; privacy, anonymity and confidentiality; and data quality. We illustrate methods of addressing these issues with our experiences conducting online surveys in educational contexts. Moving beyond the procedural ethics approach commonly adopted in quantitative educational research, we recommend adopting a situated/process ethics approach to identify and respond to ethical issues that may arise during the conduct, analysis and reporting of online survey research. The benefits of online surveying in comparison to traditional survey methods are highlighted, including the potential for online surveys to provide ethically defensible methods of conducting research that would not be feasible in offline education research settings.

Keywords: Research ethics; online survey, internet survey, higher education.

Exploring ethical issues associated with using online surveys in educational research

In the 20+ years since NCSA Mosaic provided the first major graphical interface for the world-wide-web (National Centre for Supercomputing Applications, NCSA, 2013), global Internet penetration has rapidly increased. By the end of 2014, the Internet Telecommunications Union (ITU, 2014) projects that almost 3 billion people (or some 40% of the world's population) will be defined as 'Internet users', although access remains skewed in favour of developed nations, as well as the wealthy, educated and young (Australian Bureau of Statistics, ABS, 2014). As Internet penetration has increased, researchers have been quick to identify the data collection opportunities it affords (Lee, Fielding, & Blank, 2008; Reips, 2007, 2012; Skitka & Sargis, 2006). Currently, a diverse range of research methods are employed online (including qualitative, observational/non-reactive and experimental methods), although web surveying dominates (Buchanan & Hvizdak, 2009; Krantz & Williams, 2010; Reips, 2012). Its popularity can be linked to the advantages it offers over traditional (offline) methods, including cheap, flexible, rapid access to large, diverse, geographically disparate and otherwise difficult to access samples, reduced social desirability and experimenter expectancy effects, and the ability to impose complex conditional logic on the presentation of items and stimuli (Best & Krueger, 2004; Evans & Mathur, 2005; Gosling, Vazire, Srivastava, & John, 2004; Hewson & Laurent, 2008; Skitka & Sargis, 2006; Tuten, 2010). However, web surveying is not without its challenges, including reduced experimenter control (Stieger & Reips, 2010), relatively low response rates (Shih & Fan, 2008), relatively high levels of item non-response (Heerwegh & Loosveldt, 2008) and dropout (Peytchev, 2009), and some unique ethical considerations which require addressing (Allen & Roberts, 2010; Buchanan & Williams, 2010). It is these ethical considerations, and how they relate to the conduct of online surveys in educational research, which form the substance of the current paper.

When we refer to educational research, we are referring to research “concerned with investigating all aspects of the education world” (Regan, Baldwin, & Peters, 2012, p. 45), and not just the pedagogic research (undertaken by teachers to investigate the efficacy of their work within their own schools or classrooms, Stierer & Antoniou, 2004) that it subsumes. Furthermore, we are focusing our review on the ethical use of online surveying in educational research in higher education contexts, where the potential research participants, most commonly students and/or staff, are almost always adults, and engaged with an educational institution of their own volition. The use of online surveys with children within primary and secondary education raises a host of additional ethical issues, which are beyond the scope of this paper.

The use of online surveys in educational research has grown rapidly over the past 10 years, where they have been used to shed light on topics as diverse as student evaluations of teaching (Berk, 2012) and modes of delivery (Evans, 2008), student attitudes toward forensic science (Horton et al., 2012), learning in virtual environments (Shea & Bidjerano, 2010), lecturers’ attitudes and beliefs about pedagogy for education for sustainable development (Cotton, Warren, Maiboroda, & Bailey, 2007) and professional learning of higher education teachers (Knight, Tait, & Yorke, 2006). This is unsurprising, considering the benefits of online surveys outlined above and the finding that online surveys appear preferred to paper-based surveys by both students and teachers (Roberts & Allen, 2010, 2012; Harlow, 2010). However, the ethical use of online surveys in educational research requires consideration of a range of potential issues.

Ethical issues associated with online surveys in educational research

Ethical issues associated with the use of online surveys in educational research mirror ‘generic’ ethical issues in the use of online surveys, but with an overlay of complexities resulting from the sensitivities of conducting research within educational contexts. Despite

online survey research being the most frequently reviewed type of Internet research (Buchanan & Hvizdak, 2009), not all ethics review boards may be fully cognisant of the range of ethical issues associated with online surveys generally, or as applied to educational research specifically, highlighting the need for educational researchers to be cognisant of the range of ethical issues associated with this type of research.

In this paper we describe five sets of ethical issues associated with conducting online survey research, with particular reference to their use within education research. Ethical research balances potential benefits arising from the research against potential harm to research participants or others (The National Health and Medical Research Council, the Australian Research Council and the Australian Vice-Chancellors' Committee, 2007), and we situate our discussion of the five sets of ethical issues associated with online surveys within this definition of ethical research, with particular reference to the American Educational Research Association's (AERA) 2011 *Code of Ethics*.

Dual teacher/researcher roles

As observed by Hammack (1997), teachers have a primary obligation to their students, whereas for researchers, the primary obligation is to their field of expertise. The 'good' teacher seeks to maximise opportunities for learning, whilst the 'good' researcher seeks to maximise participation, publications and data quality. When one's responsibilities as a teacher and researcher overlap, as they do when conducting research with one's own students as participants, role conflict can occur (Ferguson, Yonge, & Myrick, 2004), which is an issue often not adequately addressed in educational research proposals (Regan et al., 2012). In these situations, meeting our students' educational needs should supersede our own research needs (Brown, 2010; Regan et al., 2012), although this is often easier said than done. Indeed, on reflection, the second author can think of several instances where he has enthusiastically encouraged his students to participate in his own research, and only begrudgingly permitted

colleagues to do similar. He has mentally justified this (rather selfish) reluctance by questioning the educational merit of participation (after all, surely the students' time could be better spent studying the content of his course!). Realistically though, it probably has more to do with competition for participants, and a concern that one more participant for a colleague could mean one less for him (Adams & Umbach, 2012; Porter, Whitcomb, & Weitzer, 2004). In other words, he had been allowing his needs as a researcher to outweigh his students' educational needs, whilst arguing to himself that the opposite was true.

It is for reasons like these that the AERA (2011) Code of Ethics (Clause 14.02) specifies that educational researchers should ideally recruit participants unrelated to their other professional roles, such as teacher or supervisor. Where circumstances necessitate sampling from one's own classes, and thus dual roles do exist, the resulting research is not inherently unethical (Regan, 2013). However, the students/participants should be considered members of a 'vulnerable' population (Chen, 2011; Leentjens & Levenson, 2013) and extra care should be taken to minimise risk and adverse consequences, should students choose not to participate or withdraw consent before a study's completion. As this issue of dual-roles colours how all other ethical issues should be considered and addressed, it will be revisited several times in the following sections.

Informed voluntary consent

A basic standard of ethical research is that prospective participants are able to make informed choices about whether or not to consent to participate. Providing sufficient information to enable informed consent has been identified by Human Research Ethics Committees (HRECs) as an ethical concern across educational research proposals, with the most common ethical transgression observed by Regan (2013) in her analysis of feedback given to educational researchers being the provision of limited or incorrect information, including the use of favourably worded information to increase the likelihood of participation. A number of

Brody, Cluck, and Aragon's (1997) undergraduate student sample also reacted negatively to what they saw as "vague", "inaccurate" and "incomplete" information provided prior to participation in a diverse range of psychological studies (p. 291).

The AERA Code of Ethics (2011; Clause 13.01a) specifies that, as a general principle, educational researchers must "obtain and document written or oral consent" from research participants. However waivers of consent may apply to online surveys where the research is minimal risk and could not be practicably completed if written or oral informed consent were required (Clause 13.01b). This does not absolve researchers from fully informing potential participants about the study, extent of confidentiality, possible risks and benefits, the voluntary nature of participating and the lack of negative consequences should the individual decline to participate or withdraw from the study (Clause 13.02d).

To ensure potential participants are as fully informed in online survey research as in other types of research, Mahon (2013) recommends setting an information sheet as the first page of the online survey, with participants required to check a box to indicate consent before accessing the survey. This ensures that participants have access to the same information they would receive prior to completing an offline survey. We have used similar processes successfully in our research, hosting the information sheet on a university server and then automatically redirecting participants to an externally hosted survey (e.g., on Qualtrics.com) on consent (Roberts & Allen, 2013). To prevent prospective participants from bypassing an information sheet, many survey software systems allow the researcher to enable referrer verification (thus only allowing participants access to survey if they have come from a specific URL) and prevent search engine indexing.

Although informed consent is a necessary characteristic of ethical research, it is not a sufficient characteristic. In ethical research, consent should also be given (and withdrawn) voluntarily (e.g., AERA, 2011, Clause 14.02). When sampling from a vulnerable population,

extra care should be taken to ensure that consent is not coerced (AREA, 2011, Clause 13.01d). Although the term ‘vulnerable’ is most commonly associated with children, minorities, and individuals with special needs, the vulnerability of students in tertiary education also requires careful consideration (Chen, 2011; Leentjens & Levenson, 2013). Higher education students (in contrast to school students) are adults who have voluntarily chosen to continue their studies (Stierer & Antonious, 2004). Further, many higher education students conduct research themselves and are knowledgeable about research ethics, which potentially reduces their vulnerability (Parsell, Ambler, & Jacenyik-Trawogger, 2014). However, higher education students may still be considered a vulnerable population when the research is being conducted by a researcher with whom they also have an educational relationship (e.g., a teacher). Indeed, it has been argued that an adult student “may be competent to make decisions in general while not being competent in particular situations” (Clark & McCann, 2005, p. 44), such as when asked to participate in their own teacher’s research. In such situations, their “competence to refuse may be impaired” (p. 45).

The power imbalance between teacher and student may limit students’ abilities to freely consent where it is feared that non-participation may adversely affect their education (Ferguson, Myrick, & Yonge, 2006). When Forester and McWhorter (2005) asked 524 medical students whether or not they would feel coerced if asked to participate in faculty research, nine percent indicated they would. In a smaller sample of psychology students, Miller and Kreiner (2008) found that 25% had felt coerced or forced into participation at some point, while 33% indicated that they would feel coerced if asked to participate in their own teacher’s research. Shi (2006) also reported that a number of her students felt ‘used’ in her action research in her teacher-training class.

To minimise the coercion some students may perceive when asked to participate in their teacher’s research, various strategies have been suggested. For example, using other members

of the research team (or other non-involved academics) to recruit students; recruiting broadly (e.g., via general announcements on learning management systems and notice boards) rather than sending out personalized invitations; leaving students with time (e.g., one or more days) to decide on whether or not they wish to participate after providing them with information about a study; not collecting data during class; and ensuring the researcher with the dual role does not know who has volunteered to participate in the research (Aycock, & Currie, 2013; Clark & McCann, 2005; Comer, 2009; Ferguson et al., 2006; Regan et al., 2012; Ridley, 2009). The use of academic staff not directly involved in teaching the students should continue throughout the data collection and recording phases of the research. Educational researchers should explicitly advise students that these steps have been taken in order to counter any perceptions that research participation (or abstention) may affect grades (Ridley, 2009).

When conducting online surveys, steps can be taken to reduce perceptions of coercion associated with dual teaching/research roles. In our two studies developing and validating a measure of student perceptions of the educational value of research participation (Roberts & Allen, 2013) we employed a number of steps to ensure that students were fully informed and voluntarily chose whether or not to participate in the research. In both studies, anonymous online surveys were completed by consenting students outside of class time. In the first study, students were recruited through advertisements on student learning management system sites and announcements during lectures. Participation was voluntary and no incentive for participation was offered. Interested students were provided with a link to an online participant information sheet, and upon consenting were redirected to the online questionnaire. In the second study we recruited students through a participant pool. Students could elect to participate in this study, other studies by other researchers, or complete an alternative activity not involving research participation. Mirroring the first study, interested students were provided with a link to an online participant information sheet, which linked them to an online

questionnaire. In this instance, students completing the survey were assigned credits towards their research participation requirements.

Finally, in addition to feeling like they gave their consent freely, student participants should also feel like they can withdraw said consent freely. It appears that this may not be the case in face-to-face research, particularly when data is collected in a group setting. For example, Brody and colleagues (1987) found that student participants sometimes felt too embarrassed to ‘change their mind’ during participation, or felt that quitting would violate the terms of the consent agreement, and/or reflect poorly on them. Although many of these barriers against terminating participation are already reduced when research is conducted online, there are still additional measures that researchers can take to maximise the likelihood that students will feel genuinely free to withdraw consent, should they wish to do so. For example, in our own online survey research we typically include a statement in the participant information sheet stating that consent will only be assumed if the student/participant actually completes the survey (i.e., clicks ‘submit’ on the final page), and that anyone wishing to withdraw consent can do so by simply navigating away from the survey or closing the relevant browser window/tab.

Use of incentives

The AERA Code of Ethics (Clause 14.04) allows for offering incentives for research participation, providing they are not “excessive or inappropriate” (p. 14). However, the use of incentives in educational research needs to be carefully considered to ensure that there are no perceptions of coercion (Miller & Kreiner, 2008). Monetary incentives may infer coercion where potential student participants have limited incomes (Ridley, 2009). Further, anonymity is eroded when research participants need to demonstrate they have participated in order to obtain the incentive. In the first author’s previous research examining student perceptions of the teaching of computer-assisted qualitative data analysis (Roberts, Breen, & Symes, 2013), a

token incentive was provided to all members of the sampling pool as a way of indicating appreciation for considering participating, whether or not individual members chose to participate. All students attending scheduled research methods laboratories within the last week of semester (the sampling pool) were given a chocolate frog, prior to deciding whether or not to participate. The tutor left the room and those who chose to participate completed the anonymous online survey. Other methods of providing incentives, including entry into a prize draw, were rejected as this would alert the researchers (who also taught the students) as to who had participated, removing their anonymity.¹

A further form of incentive widely used in higher educational settings is extra credit. Lecturers who offer extra credit to their own students as an incentive for participating in their own research may reduce the trust relationship with their students (Ridley, 2009). Less contentious is the use of subject or participant pools at a faculty or school level, where students elect to participate in a range of research offered by a range of researchers as a course requirement or for extra credit, as long as alternative activities to research participation are also offered. This removes the potential for perceived coercion where the lecturer is asking the student to participate in his/her own research. In the first author's previous educational survey research (Roberts & Forman, 2014; Roberts & Povee, 2014a, 2014b; Roberts & Rajah-Kanagasabai, 2013) students were recruited through an undergraduate psychology participant pool and received research credit for participating. Participation was voluntary and students could elect to take part in other studies or complete alternative written activities.

¹ It is interesting to note that in the original version of this paper submitted for review we had included detailed information on recruitment to demonstrate how we had addressed possible concerns associated with the use of incentives and potential perceptions of coercion. On the advice of an anonymous reviewer this was removed from the paper as "The procedure is described clearly, but in perhaps a bit too much detail – e.g. it is not necessary for the reader to know ... how exactly students received the invitation".

Privacy, anonymity and confidentiality

Privacy, anonymity and confidentiality are key ethical considerations in online survey research. Educational researchers must act to minimise intrusions on the privacy of research participants (AERA, 2011, Clause 12.07) at every stage of the research process. Unsolicited online survey requests may violate the privacy of the individual to a greater extent than paper requests, as they are perceived as more intrusive (Cho & La Rose, 1999). Perceived intrusion may be greater when email requests are sent to accounts that are viewed as private. Some students view email accounts issued by an educational institution as personal property (Lefever, Dal, & Matthiasdottir, 2007), increasing the likelihood that survey requests may be seen as intrusive.

When collecting data through online surveys for educational research, researchers can minimise intrusions on privacy through only collecting identifiable information where it is specifically required for research purposes (e.g., for longitudinal studies). Where identifiable information is collected in educational research, the AERA Code of Ethics specifies that reasonable precautions must be taken to protect it during storage, delivery and electronic transfer. Of course, similar precautions should be taken with all research data, regardless of whether or not it could be used to identify individual participants.

Commercial online survey systems are increasing in functionality, with some functions potentially undermining respondent anonymity and privacy. For example, the automatic collection of Internet Protocol (IP) addresses and even geolocation data by many commercial online survey hosting sites can threaten the anonymity and privacy of respondents. An IP address is assigned to a computer or mobile device each time it connects to the Internet, providing contextual information (e.g., the geographical location of the user's Internet Service Provider) that may aid in identifying survey respondents when used in combination with time and date information (see Allen & Roberts, 2010, for a more detailed explanation). While the

legal status of IP addresses as personally identifiable information varies across countries (Buchanan & Zimmer, 2012), they should be treated in online survey research as potential identifiers. IP addresses should be stripped from the dataset, preferably before saving the data file to the researcher's computer (Barchard & Williams, 2008; Benfield & Szlemko, 2006).

Buchanan and Hvizdak (2009) reported that three-quarters of the ethics review committees they surveyed did not have a designated reviewer to examine proposals for online research, and a third did not consider evaluation of privacy and security policies of commercial online survey providers to be part of their remit. More recently, Baker's (2012) survey of Institutional Review Boards (IRBs) indicated that more than a quarter (28%) do not allow collection of IP addresses in online surveys, approximately a third (32%) allow for collection of IP addresses with conditions and 40% have no policy on IP addresses.

Unique tracking links in online surveys also undermine anonymity through providing a link between survey responses and the email address of the survey respondent. More than half of IRBs surveyed by Baker (2012) approved tracking links in online surveys, but some apply conditions for their use, such as informing potential research participants that the survey is not anonymous. Similarly, longitudinal designs that require students to provide identifying information such as student number or name in order to match respondents across time points cannot be promoted as anonymous surveys.

Even where IP addresses are not collected, tracking links are not used and identifying information is not requested within the survey itself, there is still the potential for breaches of anonymity and privacy in online surveys that are beyond the researcher's control. No online transaction can be guaranteed as completely secure due to the potential for hacking and other malicious activity. Consequently, Mahon (2013) argued that researchers should not state that online surveys are anonymous and recommended the inclusion of a warning statement in participant information sheet to that effect. While earlier research (Buchanan & Hvizdak, 2009)

suggested that some IRBs did not have a good understanding of the issues involved in online surveys or adequate processes in place to review this type of research, on the basis of a review of policy from 52 IRBs, Baker (2012) concluded that IRB policy now demonstrates sufficient understanding of these issues.

Data quality

How confident can we be that data collected in online surveys is of sufficient quality for research purposes? Obtaining quality data is an essential component of ethically defensible research, justifying the research burden placed on participants, resources consumed and investment by funders and society (Rosenthal, 1994). Further, failure to obtain quality data may result in inaccurate conclusions being drawn. Within the educational context, time spent by students, teachers and researchers on research that does not result in data of sufficient quality may be better spent on educational experiences (Rosenthal, 1994).

There are a number of factors that may limit the quality of data collected. The first of these is the representativeness of the sample obtained. If relying on email to recruit participants for educational research, some email addresses are likely to be incorrect. Lefever and colleagues (2007) reported that 8% of emails sent to students and teachers were returned as incorrect addresses. Even where email addresses are valid, they may not be accessed regularly by potential participants (Lefever et al., 2007), or invitations to participate may be erroneously filtered into a 'spam' folder. In combination, these factors may reduce the response rate to the survey and potentially bias results if unreachable potential participants systematically vary from those who do receive and read the recruitment email.

Of those who do receive and read recruitment material, not all may choose to participate (survey non-response). Survey response rates have been in decline over recent decades (Peytchev, 2013) with low response rates associated with increased sampling error and possible survey non-response bias. Survey non-response bias refers to possible differences between

respondents and non-respondents on the issues of interest (Berk, 2012), resulting in inaccurate estimates of population parameters.

In educational research, online surveys are most commonly used to capture student evaluations of teaching (SETs). In reviewing previous literature on online survey non-response on SETs, Berk (2012) noted that the responses rates for online SETs (generally around 50%) have been consistently lower than for paper-and-pencil SETs (~70-90%). Berk identified seven contributing factors to non-response in online SET surveys. Student factors were apathy, perceived lack of importance and inconvenience. Factors relating to the technology were technical problems, perceived lack of anonymity and inaccessibility. A factor relating to the survey itself was the time required for completion. Berk detailed 20 strategies that can be employed to increase response rates to online set surveys. These included the stronger marketing of SETs to students (including advertising, specifying the intended use of survey ratings, having the survey promoted by faculty, and sending of reminders); ensuring ease of access to an intuitive survey system that protects anonymity and confidentiality; offering incentives to students, faculty and departments; and in-class administration. Other strategies that may be less ethically defensible included providing students who complete SETs with earlier access to their marks, or assigning grades or extra credit to students who complete them (Berk, 2012).

A further factor that may contribute to survey non-response across all types of surveys is survey fatigue. Survey fatigue is not specific to online surveys, but a response to frequent requests to participate in survey research from a range of sources. The demand to participate in multiple surveys increases the respondent burden and results in suppressed response rates (Porter et al., 2004). Within the higher education sector the use of surveys has been increasing and includes national surveys, institutional surveys (with further surveys from individual faculties and schools) and accrediting body surveys (Porter et al., 2004). Further, within the

higher education sector, the increasing requirement for postgraduate students and academics to conduct research is resulting in increased requests to students to participate in survey research (Scott & Fonseca, 2010). Survey fatigue has been demonstrated to affect response rates to SETs, with responses rates declining once a threshold of survey requests has been received (Adams & Umbach, 2012). The ease of developing online surveys has been posited to increase the number of survey requests received and hence may further increase survey fatigue and further suppress response rates (Porter et al., 2004).

The second component of survey non-response is where participants choose not to answer some questions on a survey. Internet survey researchers can enable 'forced responding', where a participant cannot move on to a further question until an answer to the current question has been provided. While this has advantages for the researcher in terms of eliminating missing data (although perhaps just replaces it with higher rates of drop-out; see Stieger, Reips, & Voracek, 2007), it does raise ethical concerns. Baker (2012) reported that three quarters of 52 IRBs surveyed viewed forced responding as violating research participants' rights not to answer individual questions. This was particularly of concern where participation incentives were provided. Similarly Mahon (2013) argued that forced responding violates informed consent, where every research participant should be able to skip a question if they so choose. More ethically acceptable alternatives to the use of forced response validation are the use of 'decline to answer' (Baker, 2012), 'no response' or 'not applicable' options (Mahon, 2013). Further, some survey software tools enable the use of 'prompts' or 'reminders' that alert the participant to missing answers, without prohibiting continuation with the remainder of the survey.

Another factor that potentially affects the quality of data collected using online surveys is the potential for any individual to respond multiple times. Multiple responding can be detected through checking for identical IP addresses on consecutive cases and determining if

(near) identical responses have been submitted from the same IP address. Multiple responses can be deleted, retaining only the first response (Gosling et al., 2004), a practice we have used in our own research (Roberts & Allen, 2013). Participants can also be asked if they have completed the survey previously, and responses from those answering affirmatively deleted (Gosling et al., 2004). However, the risk to survey findings appears small with evidence to date suggesting multiple responding has little impact on survey findings (Gosling et al., 2004).

A further concern raised in relation to data quality is the potential for careless responding by students who perceive their participation in online surveys to be coerced (Meade & Craig, 2012). Approximately 10-12% of undergraduate students provide data that indicates careless responding is likely (Meade & Craig, 2012). A number of measures of identifying careless responding have been developed: the inclusion of items designed to detect careless responding, response consistency indices, survey response times and self-report measures of effort (Meade & Craig, 2012). The detection of careless responding enables researchers to remove such cases prior to analysis.

Given the potential threats reviewed above, how confident can we be with the results of online surveys in educational research? Reviews conducted to date suggest that equivalent results are obtained using online and offline surveys (Gosling et al., 2004; Roberts, 2006), with the possible exception of measures that may be subject to social desirability response sets (Roberts, 2006). For example, based on a 10 year longitudinal study of 63,000 student responses to SETs, Risquez, Vaughan, and Murphy (2014) reported that after controlling for class size, faculty, year of evaluation, years of teaching experience and student performance, the effect of administration mode on SET results is minimal. Further, reviewing meta-analyses of non-response bias studies, Peytchev (2013) noted that there is little evidence of a relationship between response rate and non-response bias. While steps should be taken to maximise response rates and screen data for multiple and careless responses, at the present time there is

no reason to assume that online surveys will provide lower quality data than their paper-and-pencil counterparts.

Procedural and process ethics

The five sets of ethical issues detailed above require consideration in the design phase of online survey research. Documentation of processes and procedures to be adopted based on these considerations should form part of the material prepared for formal ethics review. That is, they form part of the formal *procedural ethics* process. It is possible that during the conduct of the research project new ethical issues may emerge that were not considered as part of the procedural ethics process prior to the research commencing. Adopting a *situated ethics* approach (Simons & Usher, 2000; also known as *ethics in practice* and *process ethics*; Guillemin & Gillam, 2004) requires ethical consideration throughout the research process as events or issues arise (Guillemin & Gillam, 2004). Whilst the situated ethics approach is most strongly associated with qualitative research (Simons & Usher, 2000), it has been applied to quantitative (Jones, 2000) and online (James & Busher, 2007) education research, and in our opinion is applicable to online survey research in educational contexts. Potential ‘ethically important moments’ (Guillemin & Gillam, 2004) in online survey research requiring ethical consideration might include actual threats to data security and participant anonymity (e.g., as a result of hacking or data breaches), responding to student participants’ concerns about anonymity and/or coercion, and negotiating between conflicting interests when our teaching and research interests are not aligned. Requiring sensitivity to situational factors, educational researchers need to respond to ethical issues as they arise throughout the research, analysis and reporting process in order to minimise the potential for harm to research participants and others.

Discussion

Online surveys provide a useful tool for conducting educational research. In this paper we have outlined five areas requiring ethical consideration when using online surveys for data

collection: dual teacher/researcher roles; informed voluntary consent; use of incentives; privacy, anonymity and confidentiality; and data quality. We note that these are areas worthy of ethical consideration in all types of educational research, but require additional consideration when applied to online surveys, and advocate careful consideration of both procedural and process ethics.

These areas of ethical concern are worth addressing because online surveys provide such an efficient and flexible way of collecting data for educational research. Online surveys are preferred by both students and teachers (Roberts & Allen, 2010, 2012; Harlow, 2010) and also allow for the collection of data from students without taking up valuable (and limited) class time (Lefever et al., 2007). We have illustrated how the online survey process can be designed to offer genuine anonymity to respondents, circumventing many of the dual-role concerns that are commonly faced by educational researchers, and providing greater confidence that consent is truly voluntary, rather than influenced by perceptions of coercion.

Further, online surveys can provide ethically defensible methods of conducting research that would not be feasible in offline education research settings. Using online surveys it is possible to randomly assign research participants to conditions, creating an experiment within a survey. For example, Roberts and Rajah-Kanagasabai (2013) randomly assigned students to simulated discussion board threads that varied only in whether postings were anonymous or identified. The full potential of online surveys for education research has yet to be realised, with the rapidly increasing feature sets of online surveying tools providing an ever widening range of possibilities for survey data collection.

Future research

Ethical research is based on the premise that the potential benefits from the research outweigh potential risks to research participants and others. Student research participation provides clear benefits to educational researchers through providing access to research

participants, with possible flow-on benefits to later students through the application of findings stemming from the research. However, the benefits to participating students in terms of educational gains (including the development of “practical wisdom”; Chen, 2011, p. 281) are seldom assessed, and to our knowledge have not been assessed in relation to completion of online surveys within the educational research domain. A suggestion for future research is to examine student perceptions of the educational value of participating in educational research. The Student Perceptions of the Educational Value of Research Participation Scale (SPEVRPS, Roberts & Allen, 2013) is a brief, reliable eight-item measure that may be a useful addition to online surveys conducted within the educational context. The results would enable researchers to assess the educational value students perceive from participating in their research projects, which is a necessary first step for any teacher-researchers seeking to maximize such value. A more comprehensive evaluation of the risks and benefits to research participants could be obtained by administering this measure with the Reactions to Research Participation Questionnaire-Revised (RRPQ-R; Newman, Willard, Sinclair, & Kaloupek, 2001), which includes measures of personal benefits, emotional reactions and perceived drawbacks to participating in a study.

Conclusion

In this paper we have examined the use of online surveys in education research in relation to five key ethical issues: dual teacher/researcher roles; informed voluntary consent; use of incentives; privacy, anonymity, and confidentiality; and data quality. We have illustrated methods of addressing these issues, and recommended the adoption of a situated/process ethics approach to support traditional procedural ethics. We conclude that online surveys can provide ethically defensible methods of conducting educational research.

References

- Adams, M. J., D., & Umbach, P. D. (2012). Nonresponse and online student evaluations of Teaching: Understanding the influence of salience, fatigue and academic environments. *Research in Higher Education, 53*, 576-591. doi:10.1007/s11162-011-9240-5
- Allen, P. J., & Roberts, L. D. (2010). The ethics of outsourcing online survey research. *International Journal of Technoethics, 1*, 35-48. doi:10.4018/jte.2010070104
- American Education Research Association. (2011). *Code of ethics*. Retrieved from [http://www.aera.net/Portals/38/docs/About_AERA/CodeOfEthics\(1\).pdf](http://www.aera.net/Portals/38/docs/About_AERA/CodeOfEthics(1).pdf)
- Australian Bureau of Statistics. (2014). *Household use of information technology, Australia, 2012-13* (Catalogue Number 8146.0). Canberra, Australia: Author. Retrieved from <http://www.abs.gov.au/Ausstats/abs@.nsf/mf/8146.0>
- Aycock, D. M., & Currie, E. R. (2013). Minimizing risks for nursing students recruited for health and educational research. *Nurse Educator, 38*(2), 56-60. doi:10.1097/NNE.0b013e3182829c3a
- Baker, T. D. (2012). Confidentiality and electronic surveys: How IRBs address ethical and technical issues. *IRB: Ethics & Human Research, 34*(5), 8-15. Retrieved from <http://www.thehastingscenter.org/Publications/IRB/>
- Barchard, K. A., & Williams, J. (2008). Practical advice for conducting ethical online experiments and questionnaires for United States psychologists. *Behavior Research Methods, 40*, 1111-1128. doi:10.3758/BRM.40.1111
- Benfield, J. A., & Szlemko, W. J. (2006). Internet-based data collection: Promises and realities. *Journal of Research Practice, 2*(2), D1. Retrieved from <http://www.jrp.icaap.org/index.php/jrp/article/view/30/51>

- Berk, R. A. (2012). Top 20 strategies to increase the online response rates of student rating scales. *International Journal of Technology in Teaching and Learning*, 8(2), 98-107. Retrieved from http://www.sicet.org/journals/ijttl/issue1202/2_Berk.pdf
- Best, S. J., & Krueger, B. S. (2004). *Internet data collection*. Thousand Oaks, CA: Sage.
- Brody, J. L., Cluck, J. P., & Aragon, A. S. (1997). Participants' understanding of the process of psychological research: Informed consent. *Ethics & Behavior*, 7, 285-298. doi:10.1207/s15327019eb0704_1
- Brown, P. U. (2010). Teacher research and university institutional review boards. *Journal of Early Childhood Teacher Education*, 31, 276-283. doi:10.1080/10901027.2010.500559
- Buchanan, E. A., & Hvidzak, E. E. (2009). Online survey tools: Ethical and methodological concerns of Human Research Ethics Committees. *Journal of Empirical Research on Human Research Ethics*, 4, 37-48. doi:10.1525/jer.2009.4.2.37
- Buchanan, T., & Williams, J. E. (2010). Ethical issues in psychological research on the Internet. In S. D. Gosling & J. A. Johnson (Eds.), *Advanced methods for conducting online behavioral research* (pp. 255-272). Washington, DC: American Psychological Association.
- Buchanan, E., & Zimmer, M. (2012). Internet research ethics. *Stanford Encyclopedia of Philosophy*. Retrieved from <http://plato.stanford.edu/entries/ethics-internet-research/>
- Chen, R. P. (2011). Student participation in health professions education research: In pursuit of the Aristotelian mean. *Advances in Health Science Education*, 16, 277-286. doi:10.1007/s10459-009-9164-4
- Cho, H., & LaRose, R. (1999). Privacy issues in internet surveys. *Social Science Computer Review*, 17, 421- 434. doi:10.1177/089443939901700402

- Clark, E., & McCann, T. (2005). Researching students: An ethical dilemma. *Nurse Researcher*, *12*(3), 42-51.
- Comer, S. K. (2009). The ethics of conducting educational research on your own students. *Journal of Nursing Law*, *13*(4), 100-105. doi:10.1891/1073-7472.13.4.100
- Cotton, D. R. E., Warren, M. F., Maiboroda, O., & Bailey, I. (2007). Sustainable development, higher education and pedagogy: a study of lecturers' beliefs and attitudes. *Environmental Education Research*, *13*, 579-597. doi:10.1080/13504620701659061
- Evans, C. (2008). The effectiveness of m-learning in the form of podcast revision lectures in higher education. *Computers & Education*, *50*, 491-498. doi:10.1016/j.compedu.2007.09.016
- Evans, J. R., & Mathur, A. (2005). The value of online surveys. *Internet Research*, *15*, 195-219. doi:10.1108/10662240510590360
- Ferguson, L.M., Myrick, F., & Yonge, O. (2006). Ethically involving students in faculty research. *Nurse Education Today*, *26*, 705-711. doi:10.1016/j.nedt.2006.07.021
- Ferguson, L. M., Yonge O., & Myrick, F. (2004). Students' involvement in faculty research: Ethical and methodological issues. *International Journal of Qualitative Methods*, *3*(4), Article 5. Retrieved from http://www.ualberta.ca/~iiqm/backissues/3_4/html/ferguson.html
- Forester, J. P., & McWhorter, D. L. (2005). Medical students' perceptions of medical education research and their roles as participants. *Academic Medicine*, *80*, 780-785.
- Gosling, S. D., Vazire, S., Srivastava, S., & John, O. P. (2004). Should we trust web-based studies? A comparative analysis of six preconceptions about internet questionnaires. *American Psychologist*, *59*, 93-104. doi:10.1037/0003-066X.59.2.93
- Guillemin, M., & Gillam, L. (2004). Ethics, reflexivity, and "ethically important moments" in research. *Qualitative Inquiry*, *10*, 261-280. doi:10.1177/1077800403262360

- Hammack, F. M. (1997). Ethical issues in teacher research. *Teachers College Record*, 99, 247-265.
- Harlow, A. (2010). Online surveys-possibilities, pitfalls and practicalities: The experience of the TELA evaluation. *Waikato Journal of Education*, 15, 95-108. Retrieved from <http://researchcommons.waikato.ac.nz/bitstream/handle/10289/6163/Harlow%20online.pdf>
- Heerwegh, D., & Loosveldt, G. (2008). Face-to-face versus web surveying in a high-internet-coverage population: Differences in response quality. *Public Opinion Quarterly*, 72, 836-846. doi:10.1093/poq/nfn045
- Hewson, C., & Laurent, D. (2008). Research design and tools for Internet research. In N. Fielding, R. M. Lee, & G. Blank (Eds.), *The SAGE handbook of online research methods* (pp. 58-78). London, England: Sage.
- Horton, R. C., Kelly, T. L., Lenehan, C. E., Lennard, C., Lewis, S. W., Lim, K. F., ... & Southam, D. C. (2012). Assessing students' attitudes toward forensic science: collecting an expert consensus. *Forensic Science Policy & Management: An International Journal*, 3, 180-188. doi: 10.1080/19409044.2013.849780
- Internet Telecommunications Union. (2014). *The world in 2014: ICT facts and figures*. Retrieved from <http://www.itu.int/en/ITU-D/Statistics/Documents/facts/ICTFactsFigures2014-e.pdf>
- James, N., & Busher, H. (2007). Ethical issues in online educational research: Protecting privacy, establishing authenticity in email interviewing. *International Journal of Research & Method in Education*, 30, 101-113. doi:10.1080/17437270701207868
- Knight, P., Tait, J., & Yorke, M. (2006). The professional learning of teachers in higher education. *Studies in Higher Education*, 31, 319-339. doi:10.1080/03075070600680786

- Jones, K. (2000). A regrettable oversight or a significant omission? Ethical considerations in quantitative research in education. In H. Simons & R. Usher (Eds.), *Situated ethics in education research* (pp. 147-161). New York, NY: Routledge/Falmer.
- Krantz, J. H., & Williams, J. E. (2010). Using graphics, photographs and dynamic media. In S. D. Gosling & J. A. Johnson (Eds.), *Advanced methods for conducting online behavioral research* (pp. 45-62). Washington, DC: American Psychological Association.
- Lee, R. M., Fielding, N., & Blank, G. (2008). The Internet as a research medium: An editorial introduction to 'The Sage Handbook of Online Research Methods'. In N. Fielding, R. M. Lee, & G. Blank (Eds.), *The SAGE handbook of online research methods* (pp. 3-20). London, England: Sage.
- Leentjens, A. F. G., & Levenson, J. L. (2013). Ethical issues concerning the recruitment of university students as research subjects. *Journal of Psychosomatic Research, 75*, 394-398. doi:10.1016/j.jpsychores.2013.03.007
- Lefever, S., Dal, M., & Matthiasdottir, A. (2007). Online data collection in academic research: Advantages and limitations. *British Journal of Educational Technology, 38*, 574-582. doi:10.1111/j.1467-8535.2006.00638.x
- Mahon, P. Y. (2013). Internet research and ethics: Transformative issues in nursing education research. *Journal of Professional Nursing, 30*, 124-129. doi:10.1016/j.profnurs.2013.06.007
- Meade, A. W., & Craig, S. B. (2012). Identifying careless responses in survey data. *Psychological Methods, 17*, 437-455. doi:10.1037/a0028085
- Miller, W. E., & Kreiner, D. S. (2008). Student perception or coercion to participate in psychological research. *North American Journal of Psychology, 10*, 53-64

- National Centre of Supercomputing Applications, University of Illinois and Urbana-Champaign (2013, December 3). *NCSA Mosaic*. Retrieved from <http://www.ncsa.illinois.edu/enabling/mosaic>
- Newman, E., Willard, T., Sinclair, R., & Kaloupek, D. (2001). Empirically supported ethical research practice: The costs and benefits of research from the participants' view. *Accountability in Research: Policies and Quality Assurance*, 8, 309-329. doi:10.1080/08989620108573983
- Parsell, M., Ambler, T., & Jacenyik-Trawoger, C. (2014). Ethics in higher education research. *Studies in Higher Education*, 39, 166-179. doi:10.1080/03075079.2011.647766
- Peytchev, A. (2009). Survey breakoff. *Public Opinion Quarterly*, 73, 74-97. doi:10.1093/poq/nfp014
- Peytchev, A. (2013). Consequence of survey nonresponse. *The Annals of the American Academy of Political and Social Science*, 645, 88-111. doi:10.1177/0002716212461748
- Porter, S. R., Whitcomb, M. E., & Weitzer, W. H. (2004). Multiple surveys of students and survey fatigue. *New Directions for Institutional Research*, 21, 63-73.
- Regan, J. -A. (2013). Risks to informed consent in pedagogic research. *Journal of Perspectives in Applied Academic Practice*, 1(1), 25-29.
- Regan, J. -A., Baldwin, M. A., & Peters, L. (2012). Ethical issues in pedagogic research. *Journal of Pedagogic Development*, 2(3), 44-54. Retrieved from <http://www.beds.ac.uk/jpd/journal-of-pedagogic-development-volume-2-issue-3/ethical-issues-in-pedagogical-research>
- Reips, U. -D. (2007). The methodology of Internet-based experiments. In A. Joinson, K. McKenna, T. Postmes, & U. -D. Reips (Eds.), *The Oxford handbook of Internet psychology* (pp. 373-390). New York, NY: Oxford University Press.

- Reips, U. –D. (2012). Using the Internet to collect data. In H. Cooper, P. M. Camic, D. L. Long, A. T. Panter, D. Rindskopf, & K. J. Sher (Eds.), *APA handbook of research methods in psychology, Vol 2: Research designs: Quantitative, qualitative, neuropsychological, and biological* (pp. 291-310). Washington, DC: American Psychological Association. doi:10.1037/13620-017
- Ridley, R. T. (2009). Assuring ethical treatment of students as research participants. *Journal of Nursing Education, 48*, 537-541. doi:10.3928/01484834-20090610-08
- Risquez, A., Vaughan, E., & Murphy, M. (2014): Online student evaluations of teaching: what are we sacrificing for the affordances of technology? *Assessment & Evaluation in Higher Education, 1-16*. doi:10.1080/02602938.2014.890695
- Roberts, L. D. (2006). Equivalence of electronic and offline measures. In R. A. Reynolds, R. Woods & J. D. Baker (Eds.), *Handbook of research on electronic surveys and measurements* (pp. 96-102). Hershey, PA: Idea Group Inc.
- Roberts, L. D., & Allen, P. J. (2010, July). *Enhancing the teaching of research methods: Student perspectives on the value of research participation*. Paper presented at the 4th International Conference on Psychology Education, University of New South Wales, Australia.
- Roberts, L. D., & Allen, P. J. (2012). Student perspectives on the value of research participation. In S. McCarthy, K. L. Dickson, J. Cranney, A. Trapp, & V. Karandashev (Eds.), *Teaching psychology around the world* (Vol 3; pp. 198-211). UK: Cambridge Scholars Publishing.
- Roberts, L. D., & Allen, P. J. (2013). A brief measure of student perceptions of the educational value of research participation. *Australian Journal of Psychology, 65*, 22-29. doi:10.1111/ajpy.12007

- Roberts, L. D., Breen, L. J., & Symes, M. (2013). Teaching computer-assisted qualitative data analysis to a large cohort of undergraduate students. *International Journal of Research & Method in Education*, *36*, 279-294. doi:10.1080/1743727X.2013.804501
- Roberts, L. D. & Forman, D. (2014, early online). Interprofessional education for first year psychology students: Career plans, perceived relevance and attitudes. *Journal of Interprofessional Care*.
- Roberts, L. D., & Povee, K. (2014a). A brief measure of attitudes towards mixed methods research in psychology. *Frontiers in Psychology*, *Article 5*,1312, doi:10.3389/fpsyg.2014.01312
- Roberts, L. D. & Povee, C (2014b). A brief measure of attitudes toward qualitative research in psychology. *Australian Journal of Psychology*, *66*, 249-256. doi:10.1111/ajpy.12059
- Roberts, L. D., & Rajah-Kanagasabai, C. (2013). "I'd be so much more comfortable posting anonymously": Identified versus anonymous participation in student discussion boards. *Australasian Journal of Educational Technology*, *29*, 612-625. Retrieved from <http://ascilite.org.au/ajet/submission/index.php/AJET/article/view/452>
- Rosenthal, R. (1994). Science and ethics in conducting, analysing and reporting psychological research. *Psychological Science*, *5*, 127-134. doi:10.1111/j.1467-9280.1994.tb00646.x
- Scott, C. L., & Fonseca, L. (2010). Overstepping the mark: Ethics procedures, risky research and education researchers. *International Journal of Research and Method in Education*, *33*, 287-300. doi:10.1080/1743727X.2010.511710
- Shea, P., & Bidjerano, T. (2010). Learning presence: Towards a theory of self-efficacy, self-regulation, and the development of a communities of inquiry in online and blended learning environments. *Computers & Education*, *55*, 1721-1731. doi:10.1016/j.compedu.2010.07.017

- Shi, L. (2006). Students as research participants or as learners? *Journal of Academic Ethics, 4*, 205-220. doi: 10.1007/s10805-006-9028-y
- Shih, T. -H., & Fan, X. (2008). Comparing response rates from web and mail surveys: A meta-analysis. *Field Methods, 20*, 249-271. doi:10.1177/1525822X08317085
- Simons, H., & Usher, R. (Eds.). (2000). *Situated ethics in education research*. New York, NY: Routledge/Falmer.
- Skitka, L. J., & Sargis, E. G. (2006). The Internet as psychological laboratory. *Annual Review of Psychology, 57*, 529-555. doi:10.1146/annurev.psych.57.102904.190048
- Stieger, S., & Reips, U. -D. (2010). What are participants doing while filling in an online questionnaire: A paradata collection tool and an empirical study. *Computers in Human Behavior, 26*, 1488-1495. doi:10.1016/j.chb.2010.05.013
- Stieger, S., Reips, U. -D., & Voracek, M. (2007). Forced-response in online surveys: Bias from reactance and an increase in sex-specific dropout. *Journal of the American Society for Information Science and Technology, 58*, 1653-1660. doi:10.1002/asi.20651
- Stierer, B., & Antonious, M. (2004). Are there distinctive methodologies for pedagogic research in higher education? *Teaching in Higher Education, 9*, 275-285. doi:10.1080/1356251042000216606
- The National Health and Medical Research Council, the Australian Research Council and the Australian Vice-Chancellors' Committee. (2007). *National statement on ethical conduct in human research (updated December 2013)*. Canberra, Australia: Author.
- Tuten, T. L. (2010). Conducting online surveys. In S. D. Gosling & J. A. Johnson (Eds.), *Advanced methods for conducting online behavioral research* (pp. 179-192). Washington, DC: American Psychological Association.

Paper 8: Difficult Decisions: A Qualitative Exploration of the Statistical Decision Making Process from the Perspectives of Psychology Students and Academics

Allen, P. J., Dorozenko, K. P., & Roberts, L. D. (2016) Difficult decisions: A qualitative exploration of the statistical decision making process from the perspectives of psychology students and academics. *Frontiers in Psychology*, 7, Article 188. doi:10.3389/fpsyg.2016.00188



Difficult Decisions: A Qualitative Exploration of the Statistical Decision Making Process from the Perspectives of Psychology Students and Academics

Peter J. Allen^{1*}, Kate P. Dorozenko² and Lynne D. Roberts¹

¹ School of Psychology and Speech Pathology, Curtin University, Perth, WA, Australia, ² School of Occupational Therapy and Social Work, Curtin University, Perth, WA, Australia

OPEN ACCESS

Edited by:

Michael S. Dempsey,
Boston University Medical Center,
USA

Reviewed by:

Rink Hoekstra,
University of Groningen, Netherlands
C. Dominik Güss,
University of North Florida, USA

*Correspondence:

Peter J. Allen
p.allen@curtin.edu.au

Specialty section:

This article was submitted to
Educational Psychology,
a section of the journal
Frontiers in Psychology

Received: 14 September 2015

Accepted: 31 January 2016

Published: 16 February 2016

Citation:

Allen PJ, Dorozenko KP and
Roberts LD (2016) Difficult Decisions:
A Qualitative Exploration of the
Statistical Decision Making Process
from the Perspectives of Psychology
Students and Academics.
Front. Psychol. 7:188.
doi: 10.3389/fpsyg.2016.00188

Quantitative research methods are essential to the development of professional competence in psychology. They are also an area of weakness for many students. In particular, students are known to struggle with the skill of selecting quantitative analytical strategies appropriate for common research questions, hypotheses and data types. To begin understanding this apparent deficit, we presented nine psychology undergraduates (who had all completed at least one quantitative methods course) with brief research vignettes, and asked them to explicate the process they would follow to identify an appropriate statistical technique for each. Thematic analysis revealed that all participants found this task challenging, and even those who had completed several research methods courses struggled to articulate how they would approach the vignettes on more than a very superficial and intuitive level. While some students recognized that there is a systematic decision making process that can be followed, none could describe it clearly or completely. We then presented the same vignettes to 10 psychology academics with particular expertise in conducting research and/or research methods instruction. Predictably, these “experts” 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 pays consideration to 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. This sensitivity was reflected in their pedagogic practices. When asked to consider the format and features of an aid that could facilitate the statistical decision making process, both groups expressed a preference for an accessible, comprehensive and reputable resource that follows a basic decision tree logic. For the academics in particular, this aid should function as a teaching tool, which engages the user with each choice-point in the decision making process, rather than simply providing an “answer.” Based on these findings, we offer suggestions for tools and strategies that could be deployed in the research methods classroom to facilitate and strengthen students’ statistical decision making abilities.

Keywords: statistics, research methods, decision making, selection skills, StatHand, decision tree, graphic organizer, teaching and learning

INTRODUCTION

Quantitative research methods have played a central role in the progress of modern psychology (Benjamin, 2014), and a knowledge of quantitative methods is recognized as essential to the development of psychological literacy (McGovern et al., 2010) and the professional competence of psychology graduates. These points are reflected in the core competencies and graduate attributes specified by accrediting agencies worldwide (e.g., American Psychological Association Board of Educational Affairs Task Force on Psychology Major Competencies, 2013; Australian Psychology Accreditation Council, 2014; British Psychological Society, 2015), and by the prominent position that quantitative methods hold in undergraduate psychology curricula (Perlman and McCann, 1999). This prominence reflects a widely held understanding that an ability to critically evaluate relevant research literature, the vast majority of which is quantitative (Kidd, 2002), is a necessary precursor to evidence-based practice (American Psychological Association Presidential Task Force on Evidence Based Practice, 2006). Engaging students regularly in all aspects of the research process is recognized as fundamental to teaching quantitative methods successfully (Bradstreet, 1996; Stoloff et al., 2015), hence the typical undergraduate psychology degree provides students with multiple opportunities to conduct empirical research, either individually or in collaboration with others (Perlman and McCann, 2005).

Selecting Appropriate Statistics

Despite their prominence and utility, quantitative research methods, and particularly statistics, are known areas of weakness for many psychology students (Garfield and Ben-Zvi, 2007; Murtonen et al., 2008). Students are known to particularly struggle with the development of “selection skills” (Ware and Chastain, 1989, p. 222), or the selection of appropriate statistical tests and procedures for different types of research questions, hypotheses and data types. For example, when Gardner and Hudson (1999) asked students to identify appropriate statistical analyses for a series of brief research vignettes, most found the task extremely difficult, and performed poorly. Even though most had completed at least six research methods and statistics units¹, they managed to identify appropriate statistics for just 25.3% of the scenarios. Gardner and Hudson coded an additional 15.7% of the students’ answers as “partially correct.” When the researchers questioned the students about how they made their decisions, several explanations for the poor performance emerged. These explanations included students misinterpreting the research scenarios, being unable to actually name known procedures, misidentifying variables’ levels of measurement, and answering based on misleading key words and tables of data (which were formatted horizontally rather than vertically, as they would typically appear in a spreadsheet).

If students are required to simply recognize, rather than recall appropriate statistics, their performance is similarly limited. For

¹In the Australian context, a “unit” refers to a single subject, typically taken alongside two or three others over a semester. The term is analogous to “course” in United States higher education parlance.

example, Ware and Chastain (1989) developed a short multiple-choice selection skill test containing questions pitched at a level they believed a typical student would be able to answer on completion of an introductory statistics unit. However, when they gave the test to students at the conclusion of such a unit, the students answered fewer than 45% of the items correctly. The researchers attributed this poor performance, at least partially, to a curriculum that presented statistical techniques “one at a time” (p. 226), and provided students with few opportunities to practice selection skills. Several other researchers have made similar observations, noting that the typical research methods and statistics unit places far greater emphasis on using known statistical techniques than it does on exploring the circumstances in which they are appropriate (e.g., Bradstreet, 1996; Quilici and Mayer, 1996, 2002; Lovett and Greenhouse, 2000; Yan and Lavigne, 2014). In other words, the difficulties that students experience when placed in situations where they must work out *which technique* to use may be simply attributable to a lack of practice.

When students are provided with opportunities to practice their selection skills, performance increases somewhat (e.g., Ware and Chastain, 1991). For example, when Quilici and Mayer (2002) trained students to focus on the structural features of research scenarios (e.g., the nature of the independent and dependent variables, and the relationship between them), rather than their surface-level characteristics (e.g., the topic of the research), their ability to correctly categorize basic scenarios according to how they would be analyzed improved. The training also improved students’ abilities to produce new scenarios with the same structural features as existing ones. However, performance was still far from perfect on both outcome measures. More recently, similar findings were reported by Yan and Lavigne (2014), who also focused their training and categorization tasks on just three basic statistical tests (i.e., independent samples *t*-test, chi-square test of contingencies, and Pearson’s product moment correlation coefficient).

These findings suggest that selection skills are underpinned by a “structural awareness” (Quilici and Mayer, 2002, p. 326), which reflects an ability to disregard the surface features of a research scenario, and instead focus on its structural features and the relations between them. Consider the following section of research vignette four, presented in Appendix A in Supplementary Material:

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. In this case, it requires firstly recognizing 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. It secondly involves constructing a generic conceptual model in which the relationships between structural features are represented. In this instance, the intent of the researcher is to use a combination of several independent variables to predict scores on a continuous dependent variable. Finally, it requires integrating the conceptual model with existing knowledge to find possible solutions. For many research scenarios there are a range of statistical techniques that could be used to analyze the data, requiring the researcher to compare possible techniques to determine the most appropriate statistical technique for the particular set of circumstances. While sometimes there may be two or more equally suitable techniques, 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 testing of assumptions in order to make the final decision.

Without assistance, students find the process described above very challenging. However, "experts" do not. While the point of transition from novice to expert in this specific context is not known, it appears to necessitate a substantial amount of experience. For example, Rabinowitz and Hogan (2008) recruited graduate students enrolled in Masters and PhD courses at a university with "a very well established psychometrics program" (p. 401) to complete a series of triad judgment tasks. In these tasks they 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, even those participants with the greatest amount of experience (i.e., those who had completed between four and eight statistics units previously) did not reliably choose on the basis of structure. Those with the least experience chose 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 these "experienced" participants reliably chose 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, Masters and PhD courses in psychology and education) rarely answered more than 50% of the scenarios they were exposed to correctly.

Beyond the focus on surface and structural components of research scenarios, little is known about how students and experts select statistical tests. The first aim of this research was to develop a rich account of the strategies that psychology students and psychology academics (with expertise in research and/or research methods instruction) use to decide which statistical tests and procedures are appropriate for different research questions, hypotheses and data types.

Decision Making Aids

The preceding section 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 such decisions incorrectly can carry substantial negative consequences. At a very 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 or thinking (Bradstreet, 1996; Chance, 2002). Collectively, these points suggest a need for aids or resources that students can rely on to facilitate the statistical decision making process, and perhaps also speed their transition from novice to autonomous expert.

Numerous such aids have been developed, including tip sheets which sort statistical tests according to their defining characteristics (e.g., Twycross and Shields, 2004), and charts which link common research goals to corresponding statistics (e.g., Beitz, 1998). However, the aids which have gained most traction are based around the idea of a "decision tree" or "graphic organizer." Such resources facilitate the decision making process by prompting the user to engage with each structural feature of their research design, as well as the hierarchical and vertical relationships between them (Schau and 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 that a correct decision will ultimately be made. In the longer term, decision trees help users integrate their knowledge of statistical concepts into coherent and organized schemata, which can be quickly and effectively activated when required (Yin, 2012).

Graphic organizers to guide statistical decision making have been used for at least half a century (e.g., Siegel, 1956; Mock, 1972), and are now commonly included in statistics textbooks (e.g., Field, 2013; Tabachnick and Fidell, 2013; Allen et al., 2014). Their inclusion in such books is supported empirically by research on the efficacy graphic organizers generally (e.g., Nesbit and Adesope, 2006) and in the context of statistical decision making specifically. For example, Carlson and colleagues (Carlson et al., 2005; Protsman and Carlson, 2008) demonstrated that graphic organizers could facilitate significantly faster and more accurate (by a multiple of three) statistical decision making, compared to more traditional methods of statistical test selection (e.g., by searching through a familiar textbook). The graphic organizer method was also significantly more popular than the textbook method amongst students.

Regardless of their popularity, traditional statistics decision trees also have a number of limitations. For example, they are often constrained by the requirement that they fit within the pages of a textbook, and when given to students without accompanying resources (e.g., definitions of key terms) they can be of limited use. Koch and Gobell (1999) attempted to overcome this limitation by translating and elaborating a paper-based decision tree for delivery on the world-wide-web. In doing so, they were able to provide students with a range of additional resources, including definitions and information about how to

run and interpret the tests that their online decision tree helped students identify. Like Carlson and colleagues (Carlson et al., 2005; Protsman and Carlson, 2008), Koch and Gobell found that students using their decision tree were better able to identify appropriate statistical tests than students in a comparison condition. Unfortunately, Koch and Gobell's website is no longer active, and many of the online statistical decision trees currently available are of dubious quality or offer little more than could be contained within a traditional paper decision tree.

Aids or resources developed for students to facilitate the statistical decision making process are most likely to be promoted by instructors (experts) and adopted by students if they are developed with expressed needs and preferences of both stakeholder groups in mind. We could locate no research that asked about such needs and preferences regarding statistical decision making aids. Therefore, the second aim of our study was to elicit students' and academics' views on the nature of resources that could facilitate the statistical decision making process.

The Current Study

As noted previously, the two key aims of the current study were to (a) develop a rich account of the strategies that psychology students and psychology academics (with expertise in research and/or research methods instruction) use to decide which statistical tests and procedures are appropriate for different research questions, hypotheses and data types; and (b) elicit students' and academics' views on the nature of resources that could facilitate the statistical decision making process. The study was conducted in two phases. In phase one, undergraduate psychology students were engaged in semi-structured interviews centered on the role and value of statistics, the process of statistical test selection, and the possible characteristics of aids which may facilitate this process. The interpretations from phase one informed the development of phase two. In phase two, psychology academics were engaged in similar interviews, which also queried their perspectives on the challenges students experience when choosing between statistical tests. The findings from both phases will be integrated in the discussion.

This research complies with the guidelines for the conduct of research involving human participants, as published by the Australian National Health and Medical Research Council (National Health and Medical Research Council, Australian Research Council and Australian Vice-Chancellors' Committee, 2007). Prior to recruitment of participants, the study was reviewed and approved by the Human Research Ethics Committee at Curtin University.

PHASE ONE: STUDENTS' DECISION MAKING

Methods

Participants

The phase one participants were nine undergraduate psychology students (five female) with a mean age of 22 years. All had recently completed one or more quantitative research methods and statistics units (median = 3; range = 1–5) and were, on

average, in their third year of study. During the interviews, participants were asked to recall their grades for each completed unit, which they did with varying levels of certainty and specificity. When aggregated, these self-reports suggest that the majority of student participants typically achieved "distinction" level grades, with the remainder averaging at the "credit" level². They were recruited via posters placed around university campuses and snowballing.

Materials and Procedure

Data were collected through semi-structured interviews conducted by a research assistant, and guided by a protocol which began by asking participants about the nature of the research methods and statistics units they had taken, and their reflections on those units. They were then directed to a set of brief research vignettes (reproduced in Appendix A in Supplementary Material), prompted to imagine they were the researcher depicted in each, and asked to describe how they would determine appropriate statistics to use. Note that participants were not asked to actually identify a test or procedure (although many did), but rather describe the process or processes they would use to identify one. Following exploration of the vignettes, participants were asked to articulate the reasoning behind the processes they described, and identify processes that others may use in similar situations. Participants were then invited to describe their previous experiences with scenarios like those presented in the vignettes, and prompted to consider the role that an ability to solve such scenarios (or knowledge of an effective process for solving them) plays in a psychology graduate's repertoire of skills. Finally, the interviews concluded by asking participants to describe a tool or resource that they could use to help them approach and solve scenarios like those depicted in the vignettes. The full semi-structured interview protocol is reproduced in Appendix B in Supplementary Material.

Eight interviews were conducted face-to-face, with the final interview conducted via Skype. Each lasted between 30 and 50 min, and was audio recorded for later transcription. Prior to each interview, participants were presented with a participant information sheet, and were given the opportunity to have any questions answered. Face-to-face participants were then asked to sign a consent form, whilst the Skype participant was asked to indicate verbal consent after the consent form had been read aloud by the interviewer. At the conclusion of each interview, and before the recording device was turned off, participants were asked to verbally re-confirm consent, as recommended by Davis et al. (2004).

Data Preparation and Analysis

The audio recordings were transcribed verbatim, and the transcripts were then independently verified for accuracy. The transcripts were imported into NVivo 10, and analyzed following the stages of thematic analysis outlined by Braun and Clarke (2006). Firstly, each transcript was read and re-read, while noting down initial impressions and ideas. Following this initial

²A "credit" indicates a final mark between 60 and 69%, and a "distinction" ranges from 70 to 79%. For reference a "credit" is typically considered "average" in Australian undergraduate degrees.

familiarization stage, the data were systematically coded in a line-by-line fashion. Codes were then collated into potential themes, which were continually reviewed and refined with reference to the source data and in consultation with team members, colleagues and the research literature. In the final stages of analysis, the themes were defined, and vivid data extracts relating to each were noted for inclusion in this paper.

Findings

Several themes emerged from analysis of the student interview data. Firstly, students overwhelmingly found statistics to be challenging, yet acknowledged their importance for success in a range of different contexts. This is reflected in the theme, “*statistics are challenging, but important.*” On the whole, they found identifying appropriate statistical tests for the research vignettes particularly difficult, which resulted in embarrassment for some participants. Many struggled to describe a coherent strategy for approaching the vignettes, however some recognized that approaching them in a coherent and systematic way is possible, and tended to reflect on the utility of flow-charts and decision-trees they had encountered in their studies. These findings are captured by the themes of “*statistical selection falls outside the comfort zone,*” and “*a tenuous grasp on an elusive process.*” The students offered a variety of suggestions when prompted to consider the format and features of “*an ‘ideal’ statistical decision making aid.*” Each of these themes is elaborated on in the following sections.

Statistics are Challenging, but Important

Some students indicated that they did not expect to be taught research methods and statistics when they started their psychology degrees (“*it was a bit of a shock initially,*” “*we were so underprepared*”). Others entered the degree with negative expectations about these subjects (“*you hear about statistics before you start psychology and you hear that that’s the main reason people drop out*”). They found their early experiences with the subject matter challenging, reporting that there was a lot of “*new*” and “*difficult*” material to learn, and that they sometimes felt “*stressed,*” “*nervous,*” “*confused,*” “*overwhelmed,*” “*overloaded,*” or “*lost.*” However, they took some console from knowing that others shared these experiences:

Everyone’s in the same boat ... knowing at the very start no one knows what they are doing and everyone feeling a bit lost, it helps you feel like, ah well, I’m not the only one that is having trouble with this.

Many students reported lacking confidence in their abilities (“*I’m just useless at this side of things*”), and that they were not “*math people.*” For example, one fourth year student explained, “*I’m a words person not a numbers person, so I was really stressed about doing statistics at uni.*” One particular source of anxiety was an exaggerated concern over the consequences of making mistakes:

Having to figure out what test I was going to use ... and still thinking, okay I’m certain, but I’m also a bit unsure. If I pick the wrong test [it will have] a domino effect. Everything else isn’t going to work. It ... made me feel so nervous.

With experience, the subject matter became more manageable, and students’ confidence grew. For example, one third year student remarked that, “*once you’ve got your foot in the door you can just sort of push through and it’s easy.*” Having “*pushed through the door,*” research methods and statistics became considerably more enjoyable and rewarding:

I loved it once I understood it. But just having to go through the stress of trying to understand... getting [tutor] to explain it to me, going over the notes and trying to understand it, getting friends to explain it to me, that was very stressful and that’s the part that I just didn’t like... But once you actually get a grip on it... I love it!

Despite the challenging nature of the subject matter, students consistently acknowledged the value of research methods and statistics to the development of critical thinking (“*you can question more things, like under what circumstances did they come to that conclusion?*”), to success in their courses, and to competence as future researchers and evidence-based practitioners.

I’m excited to do honors; to do all the data analysis by myself, and I get to find out things and interpret the numbers. It’s like bringing numbers to life, so that’s exciting!

It’s important because... psychological research drives all other psychology. It’s what forms and guides what every other psychologist will do and practice... or it should do anyway.

Statistical Selection Falls Outside the Comfort Zone

Although we did not ask participants to attempt actually solving the research vignettes, this was the first instinct for many. Most found the task too difficult. They were apologetic and expressed embarrassment at being unable to successfully complete a task they felt they ought to be able to complete:

I wish I could have done a bit better for you...

[Interviewer: Do you think that being able to solve problems like these is an important skill for psychology graduates?] Of course, it’s a bit embarrassing that I can’t do it too well.

However, there was a smaller cohort who jumped straight to a statistic. Occasionally, they did so correctly. Usually though, it was with an unwarranted level of confidence. For example, when presented with a vignette depicting the relationship between two binary variables, a student mid-way through his third year of study answered, “*so it would be a paired samples t-test. Yep that’s right. Yep, pretty sure.*”

A Tenuous Grasp on an Elusive Process

When prompted to think about the process of selecting a statistic (rather than actually identifying one), students typically struggled. This was the case even for students who had completed several research methods and statistics units:

[Interviewer: So how would that help you to decide which statistical test to use?] Um see I, see I’m thinking you’d probably want to... I’m sorry. I can’t remember, sorry.

The processes they described tended to be haphazard and inefficient, and included looking for (potentially misleading) clues in the wording of the vignettes (“these scenarios are always worded in certain ways”), searching through textbooks, lecture notes (“I would probably just look at ...every single test that I’ve learned about”), the world-wide-web and previous research addressing similar research questions (“you’ve got the journals and things like...copy their methodology”). They also reported relying on memory and prior experience or the advice of friends and teachers (“you could ask your lecturers... ‘Hey, I’m doing this assignment; what do you reckon I should use?’”). Some suggested starting by entering their data into a spreadsheet, following a process of elimination, using mnemonic devices or simply guessing:

I kinda try and I guess. I don’t know, they’re never set in stone, I just kinda think like, ‘oh that’s probably that one.’

Some students did recognize that a systematic decision making process could be followed: “you go through checklists in your head.” However, none could identify every factor requiring consideration before an appropriate statistic can be identified. Most also identified irrelevant factors. For example, in the following quote, a fourth year student correctly recognized that she needs to identify the independent and dependent variables (IV and DV), as well as the number of groups being compared. However, she did not consider the measurement levels of the variables (although a nominal IV is implied by her reference to “groups”). Furthermore, she identifies causality as an issue warranting consideration. The appropriateness of causal inference is almost entirely determined by research design, and has very little to do with choice of statistic:

Figure out the variables, the IV, DV I guess. How many groups there are, and what kind of, is it a correlational relationship? ... Is it cause and effect?

Those students who recognized a process tended to refer to graphic organizers or decision trees in their statistics textbooks. They reported that such aids facilitated statistical decision making:

The tree! The wonderful tree! It is very simple, easy to use and it pretty much points you right into the analysis that you need to do.

An “Ideal” Statistical Decision Making Aid

Knowing that students find selecting appropriate statistics challenging, we asked those in our sample to explore what might make the process easier. Many turned first to their instructors, who simultaneously helped students master conceptual issues and overcome their hesitation around statistics. When prompted to think about resources they could use independently, technologically based aids were commonly considered:

If you had a website [which] just [asked] how many variables do you have? You know, how many dependent? How many independent?

What are you looking at? What are you comparing to what? And it just tells you this is the test you use.

This idea of a digital decision tree, which focuses the user on a sequence of key decision points before providing a solution was raised often. However, not all students had a preference for digital, with one remarking that she’s prefer something in a hard copy format, “because I can write into it like different things.” Other features of an “ideal” aid included simplicity, accessibility, and multiple levels of depth, as illustrated in the following quotes:

Once you’ve got the ease-of-use down and you can easily access it, and it tells you exactly what you need to do, I think that’s probably all you need really, because once you set it up you can be autonomous and you can self-direct to what you should be doing.

It would be a merge between a super simple tree diagram, but then [a] step-by-step SPSS guide book [and] behind all that a really detailed kind of book ... something that comes in three steps: simple, medium and really detailed.

Additionally, students were aware of how the content they access on the world-wide-web is of variable quality, and expressed a preference for content endorsed by recognized “experts,” such as “a psychologist...someone who knows it’s going to be useful for other psychologists,” or “some Australian government agency.” And finally, an “ideal” aid would contain engaging examples and links to other reputable resources:

Just use like real life examples... like something to do with a person and a situation, instead of saying a group of researchers want to research rats and blah blah.

If there was a way to find more resources... a way to link you with more critical approaches to some statistical tools.

Summary

In the first phase of this study, undergraduate psychology students found our discipline’s emphasis on research methods and statistics unexpected, and they approached these subjects with apprehension. They found statistics particularly challenging, but appreciated their importance to success in a range of contexts. Making statistical decisions fell outside the comfort zones of most students, which caused some embarrassment. They had a tenuous grasp on the decision making process, but recognized resources and aids that could guide them through it. When asked to consider the format and features of an “ideal” aid, they expressed a preference for an accessible, comprehensive, and reputable resource that follows a basic decision tree logic.

In the second phase of this study, we turn our attention to the statistical decision making approaches used by psychology academics with particular expertise in conducting research and/or research methods instruction. We also explore their perspectives on the challenges students face when required to choose appropriate statistical tests and procedures, as well as their thoughts about resources that could facilitate this process.

PHASE TWO: ACADEMICS' DECISION MAKING

Methods

Participants

The second phase participants were 10 psychology academics (five female) with appointment levels ranging from lecturer to professor (with a median level of senior lecturer). Six had traditional teaching and research roles, and the remainder were research focused. All were PhD qualified, research active, publishing several papers per year, and supervising research students at the level of honors and above. They predominantly identified as quantitative researchers, although some also used qualitative methods, dependent on the topic of investigation. Half had also coordinated at least one research methods and statistics unit during at least two of the preceding three years. The academic participants were recruited via individual emails, either directly from the first author's professional network, or via colleagues. They were not financially or otherwise compensated for their participation.

Materials, Procedure, Data Preparation, and Analysis

Data were collected through semi-structured interviews conducted by the second author, who did not have a dual role (e.g., as a colleague) with any of the participants. Eight were conducted face-to-face, with the remainder conducted via Skype. As in phase one, all interviews were audio-recorded, following the procedures for obtaining consent described previously. They were guided by protocols (see Appendices C,D in Supplementary Material) that began by querying the functions that statistics play in psychological research and the psychology curriculum. Participants were then directed to the set of research vignettes (presented in Appendix A in Supplementary Material), and asked to describe and explain the process they would use to identify an appropriate statistical test or procedure for each. They were then invited to describe their previous experiences with similar vignettes, and the role that being able to solve them plays in a psychology graduate's repertoire of skills. We then described to participants what we had observed when presenting the vignettes to students in phase one of the study. Specifically, we explained that most of the students struggled to articulate a coherent process, and when they attempted to solve the scenarios they tended to do so incorrectly. We then asked participants why they thought the students found this task so difficult. Finally, participants were asked to describe a tool or resource that students could use to help them approach and solve scenarios like those depicted in the vignettes. Following the interviews, the audio recordings were transcribed, and the transcriptions were analyzed using the techniques described previously.

Findings

Like the students, the academics in the sample also described the importance of statistics, both to their work and the discipline of psychology. They saw "*statistics as a tool*" (amongst several) of research. From their vantage point, the academics also reflected on the nature and value of training in statistics, which they linked primarily to the development of critical thinking and

evidence-based practice. This is captured in the theme, "*statistical training underpins competence*." When prompted to describe the factors that influence their statistical choices, the academics described a complex, nuanced and iterative "process," during which many factors warrant consideration. Some of these factors emerge from the research question and design, whilst others are linked to characteristics of the researcher and broader contextual considerations. These findings are reflected in the theme, "*decision making is a multifaceted process*." The academic participants recognized that "*students find statistical selection challenging*," and this knowledge informed their "*pedagogic practices*." Finally, they described "*an 'ideal' statistical decision making aid*" which shared many of the features identified by the students, but placed a greater emphasis on "the process" rather than "the answer." Each of these findings is elaborated in the sections that follow.

Statistics as a Tool

When asked about the role that statistics play in their work, the academics used terms such as "*central*" and "*vital*," and suggested that research would be "*pointless*" or "*nothing*" without statistics. However, despite being necessary to quantitative research, being a quantitative researcher requires much more than just knowledge of statistics. To illustrate this point, the "*statistics as a tool*" metaphor was regularly evoked. For example, "*the way I describe it to students – it's like if you're a tradie or a carpenter, then statistics are your hammer*." Furthermore, rather than assuming a primary role in the research process, statistics are subservient to the research question and design:

The important thing about research, as far as I'm concerned, is not the statistics. That's a tool that you use at the very end in order to answer the question. The important thing in my book is the questions that you're dealing with, that you develop, and the experimental designs that you then use in order to answer your questions.

In other words, the statistics "fall out" of the design, and the design is a logical consequence of the research question. Or, to quote one of the senior academics in the sample, "*we have a question, we come up with a method of testing it, and we test it and then we move on from there. We get the answer and that the answer is given to us by statistics*." It is not (or should not be) the reverse:

I don't look at it like, 'well I like this statistic, so, I'm gonna design all kinds of studies that I can use this statistic for, or this method for'. I try and look at it the other way around, which is what you're supposed to do.

Statistical Training Underpins Competence

Participants saw the role that statistics play in psychology curricula as multifaceted, and that a rigorous background in quantitative methods can distinguish the psychology graduate from graduates of other disciplines, ("*that's what makes psychologists or psychology graduates cool and different*"). While noting that statistical literacy was a necessary precondition for conducting research, they saw the primary purpose of statistical

training as tied to the competent consumption (and evaluation) of research literature and the development of critical thinking skills:

I do think it's a very central skill that they should be able to come out and go, 'Okay. Well, I can read this paper and think they've done the appropriate analysis,' and not have to rely on conclusions the authors have drawn... You're sort of critically consuming information rather than just taking what you're told.

Participants also saw training in research methods and statistics as providing a general framework for applied problem solving: “I think that approaching complex social problems in general requires you to have an understanding of multivariate and quantitative statistics. So it makes you a more informed citizen.” Furthermore, the ability to understand and evaluate research literature and solve problems were widely regarded as necessary pre-requisites for evidence based practice: “We base our profession on the scientist-practitioner model, so the evidence base is very important and statistics are really the – what we use to establish that evidence base.” However, this sentiment was not universal, with one participant commenting that, “I’m not really aware of any data which suggests that their statistical expertise is associated with better performance as a clinician... Not everyone needs as much [training in statistics and research methods].”

Despite generally recognizing their importance, some participants noted that we do not do a good job of communicating this importance to students, which may be linked to students often only appreciating the relevance of statistics and research methods in hindsight:

I don't think the reason we include them [statistics] in psych is ever made very clear to students

The feedback I get from students is often delayed... They come back a year later and say, 'thank you, I really enjoyed that. Now I understand it.' But it's a shame. I wish they would have had that eureka moment a bit earlier ...

Decision Making is a Multifaceted Process

When prompted to explicate the factors influencing analytic choices, participants described a complex, nuanced and iterative “process,” during which many issues warrant consideration:

Often there are a number of different ways to answer a question and which one's appropriate depends on the current state of the literature, obviously the data that you've collected, what it is you want to get out of it, where it's going to be published... .

This process begins with “the question” and design, followed by the nature of the variables in the study. In fact, the prevailing attitude was that, without a clear research question and intent in mind, any discussion of statistics was premature. For example, when asked about how he would respond to a student who had research ideas, but was uncertain about the appropriate statistics, one participant stated, “I would tell them that they shouldn't worry about stats; they should worry about the questions that they have, how they can operationalize the question, put it into a research

design that will give them an answer, and then we'll worry about the stats later.” However, while “jumping” into statistics too soon was regarded as poor practice, so was leaving the development of an analytic plan too long. Doing so can prove costly, as illustrated in the reflections of one senior research focused academic:

For one of the studies for my PhD I collected a load of data and then realized it actually wasn't analyzable in SPSS ... And that's where I started realizing the importance of knowing what you're doing before you start, and not collecting data and then saying, 'well, how will I analyze this?'

When developing an analytic plan, participants most commonly looked to aspects of the study. However, personal characteristics and contextual factors can also play a role in the decision making process.

Characteristics of the Study

Having a clear understanding of the purpose and design of the study as well as the number and nature of variables were recognized as essential to being able to select an appropriate statistic. For example, when presented with the second scenario in Appendix A in Supplementary Material, an experienced research methods instructor explained:

I see a between groups three level IV. And then I see a between groups two level IV. So I'm thinking a two by three factorial design. And I'm seeing this repeated measures ... So at this point I can see there's a choice between - like the way it's written implies that the dependent measure is an average over five trials. So that's a 2 x 3 between groups design. Of course, you could look at it as a three way mixed ANOVA with 'trial' as a third factor, which allows you to look at trajectories of learning. So I'm thinking if I'm writing for a journal, a learning journal, I'm pretty sure that it would be a three way mixed design. As it's presented here though it looks like a two by three between groups design.

Participants also noted that consideration should be given to alternative options in the event that analytic plans require modification due to, for example, violated assumptions. The importance of considering Type 1 and Type 2 error rates, statistical power, and the directionality of hypotheses during the decision making process were also discussed. Notably, participants actively considered viable alternatives, and weighed up the benefits and challenges associated with different decisions. This was particularly evident when discussing the mentoring of junior researchers:

Usually I will try and elicit their ideas first, and then pose some questions if I think there are other options, and ask whether they'd considered them. And if not, why not. Or if they had considered them, but decided on an alternative method, discuss why that is.

There was also a degree of tension between what could be considered “ideal,” and what is realistic or possible. As explained by one of the instructors, “there's quite a few different ways to actually do things, of varying levels of effectiveness, and depending on the resources that you have.”

Personal Factors

Participants expressed an element of personal preference when considering appropriate analytical strategies (“I’m not a fan of mixed ANOVAs. I much prefer to go through with repeated measures ANOVAs...”), although it was recognized that such an approach does not reflect “best practice.” There was also some tension between a desire to prove competence and an appreciation that the “best” technique is not necessarily the most complex:

There is something nice about really complex designs and really complex analyses that tend to stun people into thinking, ‘you know what you’re talking about!’

I tend to err on the side of you use the technique that’s appropriate, not the fanciest one. So there’s something to be said for if a t-test answers your question, use a t-test. Like there’s no need to get all fancy just for the sake of it.

Contextual Factors

It was observed by academic participants that research is not conducted in a vacuum, and that there are factors outside the researcher’s immediate control which influence the statistical decisions they make. The first of these is the intended audience: “What people need to realize is that the choice of analysis is on par with choice of audience... [and] sometimes you have to do different analyses for different audiences.” As reviewers and journal editors are frequently gatekeepers between researchers and their broader audiences, their opinions were given particular weight: “Then you get a reviewer who has their own preference on the type of statistics they would like to be used, so you have to revise it.” At times, these opinions were seen as useful, and helped shape future decision making. At other times, they could be an impediment to progress:

I was always taught that if you’re testing mediation, you should use Baron and Kenny’s model which is now, indeed, 20 years out-of-date, and there are whole books on much better ways of doing it. And the only way I came across that was when I submitted a paper with mediation and one of the reviewers said, ‘yeah, this is okay, but there’s much more sophisticated and better ways of testing that’. It put me into touch with a whole literature which I now – anytime I’m testing mediation, we use those.

And what I have experienced this last year, actually, is that I did use different statistical methods working with [a statistical consultant]... And because they were different, they were met with – reviewers didn’t like it. They didn’t like things that they didn’t know. So you’d have to explain it, and they thought that you were trying to trip them up or trick them to get something.

Participants also made regular reference to how shifting discipline practices (and what is considered “best practice”) can influence decision making. For example, one participant described how she used simple regression techniques in her PhD. Yet, if she was examining a current PhD in which the same techniques were used, she would say “no way, go back and do something much, much better.” Furthermore, although best practice guides decision making, what defines best practice is often quite opaque:

There is uncertainty ... because there’s no black and white. It’s not really that kind of field. So you might find one article that said, ‘breaking the assumption is okay under these circumstances. You can get away with it.’ And in other circumstances you can’t. So you often get contradictory messages.

The preceding quote indicates that there may be a range of “best practices,” and what is ultimately acceptable depends both on the technique applied, as well as its justification:

With my graduate students, a lot of what I’m teaching is ‘yes there are some fundamentals, but once you get beyond that it’s about being able to determine the appropriate technique for your question and your data and then be able to justify that decision knowing that you’ll send it out for review and people will disagree with you’.

Finally, beyond an aspiration toward best practice, participants also indicated a desire to avoid (or be seen to avoid) poor practice. The poor statistical practices most commonly cited centered on “fishing” for effects and their subsequent misrepresentation in published work:

If you’re just doing post hoc analysis, but pretending that it was a priori, then you get – I’ve seen it at conferences; students claiming they did a mediated moderation on one thing and then moderated mediation on the other. And you kind of go, ‘there’s no way that was a priori. You did not go into the research with that plan!’ If you do enough statistical tests and you don’t report them, and you don’t do Bonferroni corrections, then you run the risk that something is going to be significant, just because.

Students Find Statistical Selection Challenging

Aside from a small cohort of particularly capable students, it was widely recognized by the academic participants that many students find research methods and statistics challenging sections of a psychology degree. When we described the outcomes of presenting the research vignettes to the student sample, and asked academic participants why they thought the majority of students struggled with them, a range of possibilities were suggested. Some of these appeared to be attitudes or dispositions that students brought to the degree or developed over time, whereas others reflected characteristics of the teaching methods and materials commonly used in undergraduate psychology courses.

Student Characteristics

Participants perceived that the reality of a psychology degree is often inconsistent with students’ expectations on entering the course. This could be because psychology “doesn’t sound like a course that requires a lot of statistics.” They also noted that many students approach statistics with anxiety, lack confidence in their statistical abilities, are disinterested in research methods and statistics, or do not see their relevance to their future professional lives:

Students are scared of statistics. And therefore they get a bit of a mental block, I think, and convince themselves they don’t know how to answer the question.

It’s perceived as another class they don’t like, that they don’t perceive is relevant, that they don’t understand – It’s like math at school,

'when am I ever going to use this?' Because students coming in are all gong to be clinical psychologists and we know clinical psychologists never use numbers <laughs>!

Course Characteristics

Academic participants highlighted both implicit and explicit characteristics of the research methods and statistics curriculum which may hinder, rather than support students' skill development. For example, one participant described the discipline's tendency to "fetishize" statistics, and how this value is communicated to students:

"There's an element of elitism. If we make it seem really hard and difficult to get into and make it really opaque, we're shoring up this idea that stats is for the hard men and the real - we can sort the men from the boys amongst the students and also amongst everyone else of us too."

Others spoke of teaching approaches which tend to compartmentalize content, which is stripped of context when presented to students:

It was very much pigeon-holed. So it was very much this week we're talking about ANOVA; this week, we're talking about regression; this week, we're talking about something else. So there really wasn't that opportunity to make a decision about which one is which. It was just, 'this is what you're doing'.

Overwhelmingly though, participants ascribed the difficulties students have with statistical decision making to teaching methods which don't engage students in regular decision making opportunities from early in the course ("there just isn't enough exposure to that sequence of thought and planning"), and don't regularly reinforce the relevance of statistics. It was considered that both these aims could be achieved by engaging students in the full research "process." To participants, this process begins with a substantive research question, works through key issues tied to design and analysis, and concludes with clear implications or, to quote one instructor, an answer to the question, "what does this shit actually mean?"

Showing that it's not necessarily about numbers but about answering questions might help with some of the - and putting it into that context, and putting it into the context of a research problem and not a math problem - I think, it can help as well.

Answering questions of substantive interest was seen as vital. Furthermore, failing to achieve this aim may promote disinterest, disengagement, and apathy.

... as soon as it's a question that you wanna know the answer to, it's like ... it suddenly becomes relevant and important.

Pedagogic Practices

Recognizing that statistical decision making is an area that students find challenging, participants employed a number of techniques to encourage and support their efforts. This tended to occur in the context of either small-group/individual research

supervision sessions or lab group meetings. Firstly, questioning was used to guide students "through the process."

I use a lot of questioning and I'm just thinking about one student that I spoke to just last week who put point blank to me, she said, 'oh, we'll be using [multiple] regression to answer this question,' and I immediately sort of flicked it back on her and said, 'but how are you measuring your DV?' - which was dichotomous. So in asking that question, she was able to go, 'oh hang on a minute... that data is not appropriate for what I just said'.

The process involves considering design and statistical issues concurrently, and in the context of the research question or objective:

I ask them to draw out the design of an experiment, say, and they might suggest some stats at the end. And then, I ask them how that addresses the question or questions [they] want to get to.

It also involves consideration and evaluation of different options before making decisions, and collaboration and consultation is encouraged:

... try and present the different options... what are the pros and cons of each in this case, and then weigh those and come to a decision. I think you kind of need to let them go through the process.

An "Ideal" Statistical Decision Making Aid

Academic participants suggested characteristics for a tool or resource that students could make use of to independently identify appropriate statistics for various circumstances. First, the resource should be accessible (in terms of ease and cost of availability), and step users through a sequence of questions or decisions which must be addressed to arrive at an actionable outcome. Terms like "flow-chart" and "decision-tree" were used commonly.

It is a question and answer flow-chart kind of situation. Is it relationships or differences? ... how many variables; categorical or continuous? The answers to each of those questions would lead you to the correct [statistical analysis].

It seems like if there was some sort of decision tree ... It would make sense to have some sort of app or something ... easily accessible online or on your phone or whatever, where you can plug in and go through a step-by-step process.

If questions or decision points are presented sequentially, the user is forced to engage with each step in "the process" and can thus be "train[ed] ... to ask the important questions." The longer term objective of such a resource should not be reliance, but rather a transition toward greater autonomy and flexibility:

[After using the resource for a period of time, the user should ideally be able to] turn it off or turn the book over and then you give them another problem and see, well can they now - are they now able to - even if they can't get to the right answer, are they now trying to figure out? 'Well, what am I trying to do? How many groups and

what am I - what's my IV, what's my DV, do I have more than one IV, what's the level of measurement?

Participants also noted that understanding key terms (or having the ability to quickly look them up) is essential to being able to use such a resource effectively (“*you need to know what a covariate is, what the IVs and the DVs, what this actually means*”). Finally, they acknowledged that, realistically, such a resource is never going to capture all the nuances in statistical decision making, but may be useful within the broader discussion:

If you try to reduce it to a few basic principles then you're missing critical questions, like 'what is the hypothesis' and 'what is the audience'? It's really much better if it's a consultative process with an advisor and/or with other [students]. I don't think people should work independently necessarily. I think that there's a lot of virtue in consulting with people in the design phase of the project.

Summary

In this study's second phase, the academics saw statistics as one of several tools available to the researcher; a tool that is vital to the conduct of most research, but subservient to the research question and design. They acknowledged the role that statistics training plays in the development of research skills, but saw its primary role as nurturing the development of critical thinking and evidenced-based practice. The academics described choosing an appropriate statistic as a complex, nuanced, and iterative process, during which consideration should be paid to multiple contextual factors in addition to the characteristics of the study. They were sensitive to the challenges that many students experience when making statistical decisions, which they attributed partially to how research methods and statistics are commonly taught. This sensitivity was reflected in their pedagogic practices. The “ideal” statistical decision making aid the academics described shared many of the features identified by the student participants, although greater emphasis was placed on “the process” than “the answer.”

DISCUSSION

The first aim of this research was to explore the strategies that psychology students and academics use to select statistical tests. We probed these strategies in semi-structured interviews, in which participants were encouraged to discuss how they would approach each of a series of short research vignettes. Our findings indicate a number of key differences between how these two groups approach statistical decision making.

For the students in our sample, being required to make such decisions pushed them outside their comfort zones, resulting in either apologetic discomfort, or instinctual selections that were frequently incorrect. This finding is not surprising given the body of literature demonstrating that most students find statistics generally (Garfield and Ben-Zvi, 2007; Murtonen et al., 2008), and statistical decision making specifically (Ware and Chastain, 1991; Gardner and Hudson, 1999) to be difficult. Their ability to even describe the process of selecting a test was limited, and relied heavily on the use of strategies unlikely to produce optimal

outcomes. These included searching through textbooks, lecture notes, and the world-wide-web, relying on memory and prior experience, turning to the advice of friends or teachers, and looking for clues in the wording or structure of the vignettes. A number of these strategies were also suggested or displayed by the students in Gardner and Hudson's (1999) research, who were particularly prone to misinterpreting research questions, and being misled by key words and data presentations formats. Like those in Gardner and Hudson's research, the students in our sample were reasonably far into their degrees and were, on average, in their third year of study.

There were a minority of students who recognized that a systematic decision making process could be used to approach and “solve” the research vignettes. However, none were able to identify all the factors in the vignettes that would require consideration before appropriate statistics could be identified. Furthermore, these students had a tendency to also identify features of the vignettes which were irrelevant to the task at hand. Again, these findings are broadly consistent with Gardner and Hudson (1999), whose students often failed to take the nature of data (e.g., nominal, ordinal etc.) into consideration when making statistical decisions.

By way of contrast, the psychology academics described selecting appropriate statistics as a complex, nuanced and iterative process, embedded within the broader process of conducting research. They demonstrated how during statistical decision making, consideration ought to be paid to multiple contextual factors (e.g., the intended audience, prevailing discipline trends and practices etc.), in addition to the intent and design of the study itself. These experts were able to suggest appropriate statistical analyses for each vignette with ease, but were often reluctant to do so without understanding the purpose of the research, or having an opportunity to explore alternative possibilities. This behavior is suggestive of “structural awareness,” which is an ability to see past the surface features of a problem, and focus on its structural characteristics and the relations between them (Quilici and Mayer, 2002)³. It is a characteristic common to “expert” problem solvers across a wide range of specialized domains (Rabinowitz and Hogan, 2008).

Previous research suggests that structural awareness tends to develop naturally with experience (Rabinowitz and Hogan, 2008). In the Australian context, opportunities to engage in statistical decision making are limited prior to fourth year when, under individual supervision, psychology students embark on their first major research project. During this intensive research internship, expert supervisors model the statistical decision making process, and use a range of techniques to promote its

³Despite this structural awareness, the findings suggest that some psychology research academics have preferred techniques, will at times select techniques based on what they can “sell” rather than current best practice, and are reluctant to be early adopters of new techniques. This “resistance” by substantive psychological researchers to changing statistical techniques and employing new advanced statistical techniques has previously been recognized in the research literature (Sharpe, 2013). It has been attributed to a combination of a lack of awareness of new statistical developments, inadequate statistical education, the failure of journal editors to act as catalysts for change, the pressure to “publish or perish,” and fear of deviating from normative statistical practices (Sharpe, 2013).

development in students. Students in earlier years are largely reliant on lectures, laboratories, and tutorials to develop their research skills, and alternative methods of teaching statistical test selection, which are not reliant on individual supervision, are required for these years.

Our recommendation is to provide students with regular opportunities to engage in the statistical decision making process in the context of class research projects. It is widely recognized that scaffolded immersion in all aspects of the research process, from participation and/or data collection, through the development and testing of hypotheses, to the interpretation and reporting of findings, is a particularly effective way of teaching research skills (Bradstreet, 1996; Marek et al., 2004; Roberts and Allen, 2012, 2013; Earley, 2014; Stoloff et al., 2015). This point was echoed by the academic participants in the current research, who reflected on how embedding statistical decision making in a context of substantive interest, and providing opportunities to work with personally meaningful data promotes student engagement. As an example, in the first author's second year experimental methods and statistics unit, students participate in an experiment early in the semester, which forms the basis of a research report assessment. The topic varies from year to year, but typically involves studying a well established phenomenon in a contemporary context (e.g., the attractiveness stereotype on Facebook; or the Internet as a transactive memory source). In a series of class and homework exercises, students are required to develop one or two theoretically meaningful hypotheses, use the class generated data to test them, and then prepare an American Psychological Association (APA) style research report for assessment. The experiment is usually structured such that several meaningful hypotheses are possible, and testable using techniques taught in the unit (which include parametric and non-parametric tests for comparing independent and related groups). One of the key tasks in this process is the identification of an appropriate statistical test for each hypothesis. Of course, such class research projects need not be the exclusive domain of research methods and statistics units, and can also be deployed effectively to teach a wide range of subjects (e.g., Lutsky, 1986; Ragozzine, 2002).

The second aim of this research was to solicit psychology students' and academics' views on the nature of resources that could facilitate the statistical decision making process. The findings indicate that both groups support the development of a digital decision tree that is simple to use, easy to access, provides multiple levels of depth, and is endorsed by "experts." The psychology academics also stressed the need for such a resource to function as a teaching tool, which engages students with each choice-point in the decision making process, rather than simply providing an "answer." This is in contrast to some recent trends in statistics software development to automate the test selection process based on the characteristics of the user's data file (e.g., "Nonparametric Tests" in IBM SPSS; Wacharamanotham et al., 2015). In fact, such trends are antithetical to the views of the academics in our sample, who strongly believed that statistics should be considered concurrently with other design issues, and far before any data are collected.

Based partially on the findings of the current study, as well as existing literature on the efficacy of decision trees and mobile learning technologies, we have recently published StatHand (see <https://stathand.net>), a free cross-platform mobile application designed to support students through the statistical decision making process. This application, developed with the support of the Australian Government Office for Learning and Teaching, guides users through a series of annotated questions to ultimately offer them the guidance necessary to conduct a suitable statistical test, as well as interpret and report its results. A full discussion of StatHand is beyond the scope of this paper, but interested readers are referred to Allen et al. (under review). In this paper, we overview the rationale behind StatHand, describe the development process and feature set of the application, and provide guidelines for integrating its use into the research methods curriculum.

When interpreting the findings of this research, readers should give consideration to the usual caveats regarding small samples and the transferability of qualitative research findings. The nature of the task we asked of participants (i.e., to describe how they would identify a suitable statistic) also warrants some consideration. It is plausible that the apparent deftness with which the academics approached this task is at least partially a function of the nature of their work, in which we imagine they routinely practice the metacognition and self-reflection for which we probed⁴. By contrast, it is suspected that the students in the sample have less experience with such skills, and fewer daily opportunities to practice them. However, this is a matter requiring attention in future research. Future research should also focus on exploring theoretically driven strategies and resources that may facilitate the statistical decision making process, and speed up the development of selection skills and structural awareness. To date, work in this area has largely focused on involving students in concrete research projects (e.g., Kardash, 2000) or the use of decision trees (e.g., Carlson et al., 2005; and the current research). Future work should be methodologically rigorous, and based on experimental methods, rather than the non-experimental and quasi-experimental approaches so commonly utilized in teaching and learning research (Wilson-Doenges and Gurung, 2013).

In conclusion, this paper presents a qualitative exploration of the strategies psychology students and academics use to make statistical decisions. The students in our sample found this task challenging, and many struggled to describe a coherent strategy for choosing appropriate statistical tests for common research scenarios. Those who did recognize that such scenarios could be approached in a systematic fashion tended to reflect on the utility of decision trees they had encountered in their studies. Unlike the students, the academics described selecting appropriate statistics as a

⁴As kindly noted by one of our reviewers, the apparent deftness with which the academic participants were able to explore possibilities and identify suitable statistics sits in contrast with our discipline's well known difficulties when it comes to interpreting such statistics (e.g., Cohen, 1994; Hoekstra et al., 2006, 2014; McGrath, 2011; Kline, 2013).

complex, nuanced, and iterative process, embedded within the broader process of conducting research. When both groups were asked to imagine tools or resources that could facilitate the statistical decision making process, they tended to describe digital technologies based on a decision-tree framework. To the academics in particular, it was important that such resources scaffold the development of independent decision making competence, and not strip the user of the learning opportunities inherent in working through the full research process.

AUTHOR CONTRIBUTIONS

PA conceived and designed the study with the support of LR. KD conducted the second phase interviews. PA analyzed the data and lead the writing of this manuscript, both with support and contributions from LR and KD.

REFERENCES

- Allen, P., Bennett, K., and Heritage, B. (2014). *SPSS Statistics Version 22: A Practical Guide*. Melbourne, VIC: Cengage Learning.
- American Psychological Association Board of Educational Affairs Task Force on Psychology Major Competencies (2013). *APA Guidelines for the Undergraduate Psychology Major: Version 2.0*. Washington, DC: American Psychological Association Board of Educational Affairs Task Force on Psychology Major Competencies. Retrieved from <http://www.apa.org/ed/precollge/about/psymajor-guidelines.pdf>
- American Psychological Association Presidential Task Force on Evidence Based Practice (2006). Evidence-based practice in psychology. *Am. Psychol.* 61, 271–285. doi: 10.1037/0003-066X.61.4.271
- Australian Psychology Accreditation Council (2014). *Accreditation Standard for Programs of Study in Psychology Eligible for Approval Under the National Registration and Accreditation Scheme for the Health Professions: Second Consultation Draft*. Melbourne, VIC: Australian Psychology Accreditation Council. Retrieved from https://www.psychologycouncil.org.au/Assets/Files/APAC_Accreditation_Standard_for_Programs_of_Study_in_Psychology_2nd_Consultation_Draft_27_May_2014.pdf
- Beitz, J. M. (1998). Helping students learn and apply statistical analysis: a metacognitive approach. *Nurse Educ.* 23, 49–51. doi: 10.1097/00006223-199801000-00016
- Benjamin, L. T. Jr. (2014). *A Brief History of Modern Psychology, 2nd Edn*. Hoboken, NJ: Wiley Blackwell.
- Bradstreet, T. E. (1996). Teaching introductory statistics courses so that nonstatisticians experience statistical reasoning. *Am. Stat.* 50, 69–78.
- Braun, V., and Clarke, V. (2006). Using thematic analysis in psychology. *Qual. Res. Psychol.* 3, 77–101. doi: 10.1191/1478088706qp063oa
- British Psychological Society (2015). *Standards for the Accreditation of Undergraduate, Conversion and Integrated Masters Programmes in Psychology (October 2015)*. Leicester: British Psychological Society. Retrieved from http://www.bps.org.uk/system/files/Public%20files/PaCT/undergraduate_accreditation_2015_web.pdf
- Carlson, M., Protsman, L., and Tomaka, J. (2005). Graphic organizers can facilitate selection of statistical tests. Part 1 - analysis of group differences. *J. Phys. Therapy Educ.* 19, 57–66.
- Chance, B. L. (2002). Components of statistical thinking and implications for instruction and assessment. *J. Stat. Educ.* 10. Available online at: <http://www.amstat.org/publications/jse/v10n3/chance.html>
- Cohen, J. (1994). The earth is round ($p < .05$). *Am. Psychol.* 49, 997–1003. doi: 10.1037/0003-066X.49.12.997
- Davis, M., Bolding, G., Hart, G., Sherr, L., and Elford, J. (2004). Reflecting on the experiences of interviewing online: perspectives from the Internet and HIV study in London. *AIDS Care* 16, 944–952. doi: 10.1080/09540120412331292499
- Earley, M. A. (2014). A synthesis of the literature on research methods education. *Teach. High. Educ.* 19, 242–253. doi: 10.1080/13562517.2013.860105
- Field, A. (2013). *Discovering Statistics Using IBM SPSS Statistics, 4th Edn*. London: Sage.
- Gardner, P. L., and Hudson, I. (1999). University students' ability to apply statistical procedures. *J. Stat. Educ.* 7. Available online at: <http://www.amstat.org/publications/jse/secure/v7n1/gardner.cfm>
- Garfield, J., and Ben-Zvi, D. (2007). How students learn statistics revisited: a current review of research on teaching and learning statistics. *Int. Stat. Rev.* 75, 372–396. doi: 10.1111/j.1751-5823.2007.00029.x
- Hoekstra, R., Finch, S., Kiers, H. A. L., and Johnson, A. (2006). Probability as certainty: dichotomous thinking and the misuse of p values. *Psychon. Bull. Rev.* 13, 1033–1037. doi: 10.3758/BF03213921
- Hoekstra, R., Morey, R. D., Rouder, J. N., and Wagenmakers, E.-J. (2014). Robust misinterpretation of confidence intervals. *Psychon. Bull. Rev.* 21, 1157–1164. doi: 10.3758/s13423-013-0572-3
- Kardash, C. M. (2000). Evaluation of undergraduate research experience: perceptions of undergraduate interns and their faculty mentors. *J. Educ. Psychol.* 92, 191–201. doi: 10.1037/0022-0663.92.1.191
- Kidd, S. A. (2002). The role of qualitative research in psychology journals. *Psychol. Methods* 7, 126–138. doi: 10.1037/1082-989X.7.1.126
- Kline, R. B. (2013). *Beyond Significance Testing: Statistics Reform in the Behavioral Sciences, 2nd Edn*. Washington, DC: American Psychological Association. doi: 10.1037/14136-000
- Koch, C., and Gobell, J. (1999). A hypertext-based tutorial with links to the Web for teaching statistics and research methods. *Behav. Res. Methods Instrum. Comput.* 31, 7–13. doi: 10.3758/BF03207686
- Lovett, M. C., and Greenhouse, J. B. (2000). Applying cognitive theory to statistics instruction. *Am. Stat.* 54, 196–206. doi: 10.1080/00031305.2000.10474545
- Lutsky, N. (1986). Undergraduate research experience through the analysis of data sets in psychology courses. *Teach. Psychol.* 13, 119–122. doi: 10.1207/s15328023top1303_4
- Marek, P., Christopher, A. N., and Walker, B. J. (2004). Learning by doing: Research methods with a theme. *Teach. Psychol.* 31, 128–131. doi: 10.1080/03075079512331381703
- McGovern, T. V., Corey, L., Cranney, J., Dixon, W. E. Jr., Holmes, J. D., Kuebli, J. E., et al. (2010). "Psychologically literate citizens," in *Undergraduate Education in Psychology: A Blueprint for the Future of the Discipline*, ed D. F. Halpern (Washington, DC: American Psychological Association), 9–27.

ACKNOWLEDGMENTS

Support for this project has been provided by the Australian Government Office for Learning and Teaching (Grant Ref: ID13-2954). The views in this project do not necessarily reflect the views of the Australian Government Office for Learning and Teaching. This research was also supported by a grant awarded to Peter Allen by the Research and Development Committee of the School of Psychology and Speech Pathology, Faculty of Health Sciences, Curtin University. Finally, the authors would like to acknowledge James Finlay, who conducted the first phase interviews as part of a summer scholarship in 2014.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fpsyg.2016.00188>

- McGrath, R. E. (2011). *Quantitative Models in Psychology*. Washington, DC: American Psychological Association. doi: 10.1037/12316-000
- Mock, T. J. (1972). A decision tree approach to the methodological decision process. *Account. Rev.* 47, 826–829.
- Murtonen, M., Olkinuora, E., Tynjala, P., and Lehtinen, E. (2008). “Do I need research skills in working life?”: University students’ motivation and difficulties in quantitative methods courses. *High. Educ.* 56, 599–612. doi: 10.1007/s10734-008-9113-9
- National Health and Medical Research Council, Australian Research Council and Australian Vice-Chancellors’ Committee (2007). *National Statement on Ethical Conduct in Human Research 2007 (Updated May 2015)*. Canberra, ACT: National Health and Medical Research Council, Australian Research Council and Australian Vice-Chancellors’ Committee.
- Nesbit, J., and Adesope, O. (2006). Learning with concept and knowledge maps: a meta-analysis. *Rev. Educ. Res.* 76, 413–448. doi: 10.3102/00346543076003413
- Perlman, B., and McCann, L. I. (1999). The most frequently listed courses in the undergraduate psychology curriculum. *Teach. Psychol.* 26, 177–182. doi: 10.1207/S15328023TOP260303
- Perlman, B., and McCann, L. I. (2005). Undergraduate research experiences in psychology: a national study of courses and curricula. *Teach. Psychol.* 32, 5–14. doi: 10.1207/s15328023top3201_2
- Protsman, L., and Carlson, M. (2008). Graphic organizers can facilitate selection of statistical tests: Part 2 – Correlation and regression analysis. *J. Phys. Therapy Educ.* 22, 36–41.
- Quilici, J. L., and Mayer, R. E. (1996). Role of examples in how students learn to categorize statistics word problems. *Educ. Psychol.* 88, 144–161. doi: 10.1037/0022-0663.88.1.144
- Quilici, J. L., and Mayer, R. E. (2002). Teaching students to recognize structural similarities between statistics word problems. *Appl. Cogn. Psychol.* 16, 325–342. doi: 10.1002/acp.796
- Rabinowitz, M., and Hogan, T. M. (2008). Experience and problem representation in statistics. *Am. J. Psychol.* 121, 395–407. doi: 10.2307/20445474
- Ragozzine, F. (2002). SuperLab LT: evaluation and uses in teaching experimental psychology. *Teach. Psychol.* 29, 251–254. doi: 10.1207/S15328023TOP2903_13
- Roberts, L. D., and Allen, P. J. (2012). “Student perspectives on the value of research participation,” in *Teaching Psychology Around the World: Volume 3*, eds S. McCarthy, K. L. Dickson, J. Cranney, A. Trapp, and V. Karandashev (Newcastle upon Tyne: Cambridge Scholars Publishing), 198–212.
- Roberts, L. D., and Allen, P. J. (2013). A brief measure of student perceptions of the educational value of research participation. *Aust. J. Psychol.* 65, 22–29. doi: 10.1111/ajpy.12007
- Schau, C., and Mattern, N. (1997). Use of map techniques in teaching applied statistics courses. *Am. Stat.* 51, 171–175.
- Sharpe, D. (2013). Why the resistance to statistical innovations? Bridging the communication gap. *Psychol. Methods* 18, 572–582. doi: 10.1037/a0034177
- Siegel, S. (1956). *Nonparametric Statistics for the Behavioral Sciences*. New York, NY: McGraw-Hill.
- Stoloff, M. L., Good, M. R., Smith, K. L., and Brewster, J. (2015). Characteristics of programs that maximize psychology major success. *Teach. Psychol.* 42, 99–108. doi: 10.1177/0098628315569877
- Tabachnick, B. G., and Fidell, L. S. (2013). *Using Multivariate Statistics, 6th Edn*. Boston, MA: Pearson.
- Twycross, A., and Shields, L. (2004). Statistics made simple: part 4—Choosing the right statistical test. *Paediatr. Nurs.* 16:24. doi: 10.7748/paed2004.10.16.8.24.c944
- Wacharamanotham, C., Subramanian, K., Völkel, S. T., and Borchers, J. (2015). “Statsplorer: guiding novices in statistical analysis,” in *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (New York, NY: Association for Computing Machinery), 2693–2702. doi: 10.1145/2702123.2702347
- Ware, M. E., and Chastain, J. D. (1989). Computer-assisted statistical analysis: a teaching innovation? *Teach. Psychol.* 16, 222–227. doi: 10.1207/s15328023top1604_16
- Ware, M. E., and Chastain, J. D. (1991). Developing selection skills in introductory statistics. *Teach. Psychol.* 18, 219–222. doi: 10.1207/s15328023top1804_4
- Wilson-Doenges, G., and Gurung, R. A. R. (2013). Benchmarks for scholarly investigations of teaching and learning. *Aust. J. Psychol.* 65, 63–70. doi: 10.1111/ajpy.12011
- Yan, J., and Lavigne, N. C. (2014). Promoting college students’ problem understanding using schema-emphasizing worked examples. *J. Exp. Educ.* 82, 74–102. doi: 10.1080/00220973.2012.745466
- Yin, Y. (2012). Using tree diagrams as an assessment tool in statistics education. *Educ. Assessment* 17, 22–49. doi: 10.1080/10627197.2012.697850

Conflict of Interest Statement: Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2016 Allen, Dorozenko and Roberts. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Paper 9: Introducing StatHand: A Cross-Platform Mobile Application to Support Students' Statistical Decision Making

Allen, P. J., Roberts, L. D., Baughman, F. D., Loxton, N. J., Van Rooy, D., Rock, A. J., & Finlay, J. (2016). Introducing StatHand: A cross-platform mobile application to support students' statistical decision making. *Frontiers in Psychology, 7*, Article 288. doi:10.3389/fpsyg.2016.00288



Introducing StatHand: A Cross-Platform Mobile Application to Support Students' Statistical Decision Making

Peter J. Allen^{1*}, Lynne D. Roberts¹, Frank D. Baughman¹, Natalie J. Loxton², Dirk Van Rooy³, Adam J. Rock⁴ and James Finlay¹

¹ School of Psychology and Speech Pathology, Curtin University, Perth, WA, Australia, ² School of Applied Psychology, Griffith University, Brisbane, QLD, Australia, ³ Research School of Psychology, Australian National University, Canberra, ACT, Australia, ⁴ School of Behavioural, Cognitive and Social Sciences, University of New England, Armidale, NSW, Australia

OPEN ACCESS

Edited by:

Douglas Kauffman,
Boston University School of Medicine,
USA

Reviewed by:

Courtney Haines,
University of Wyoming, USA
Ricardo Tejeiro,
University of Liverpool, UK

*Correspondence:

Peter J. Allen
p.allen@curtin.edu.au

Specialty section:

This article was submitted to
Educational Psychology,
a section of the journal
Frontiers in Psychology

Received: 21 September 2015

Accepted: 14 February 2016

Published: 29 February 2016

Citation:

Allen PJ, Roberts LD, Baughman FD,
Loxton NJ, Van Rooy D, Rock AJ
and Finlay J (2016) Introducing
StatHand: A Cross-Platform Mobile
Application to Support Students'
Statistical Decision Making.
Front. Psychol. 7:288.
doi: 10.3389/fpsyg.2016.00288

Although essential to professional competence in psychology, quantitative research methods are a known area of weakness for many undergraduate psychology students. Students find selecting appropriate statistical tests and procedures for different types of research questions, hypotheses and data types particularly challenging, and these skills are not often practiced in class. Decision trees (a type of graphic organizer) are known to facilitate this decision making process, but extant trees have a number of limitations. Furthermore, emerging research suggests that mobile technologies offer many possibilities for facilitating learning. It is within this context that we have developed StatHand, a free cross-platform application designed to support students' statistical decision making. Developed with the support of the Australian Government Office for Learning and Teaching, StatHand guides users through a series of simple, annotated questions to help them identify a statistical test or procedure appropriate to their circumstances. It further offers the guidance necessary to run these tests and procedures, then interpret and report their results. In this Technology Report we will overview the rationale behind StatHand, before describing the feature set of the application. We will then provide guidelines for integrating StatHand into the research methods curriculum, before concluding by outlining our road map for the ongoing development and evaluation of StatHand.

Keywords: statistics, research methods, selection skills, decision tree, teaching and learning, mobile learning, iOS, web application

INTRODUCTION

Quantitative research methods underpin psychological literacy (McGovern et al., 2010; Cranney and Dunn, 2011; Roberts et al., 2015), and are critical to the development of professional competence in psychology. They have featured prominently in undergraduate psychology curricula since the discipline's formation (Perlman and McCann, 1999; Saville, 2008), and are reflected in the course learning outcomes and graduate attributes specified by accrediting psychology organizations worldwide. For example, the Australian Psychology Accreditation Council [APAC] (2014, p. 35) specify six graduate attributes for an undergraduate psychology program. Two of

these, (“understands the principles of scientific method and is able to apply and evaluate basic research methods in psychology” and “demonstrates the capacity to utilize logic, evidence, and psychological science to evaluate claims about, and solve problems regarding, human behavior”), require a solid and flexible understanding of research methods and statistics. The second of five learning goals for an undergraduate psychology course detailed by American Psychological Association Board of Educational Affairs Task Force on Psychology Major Competencies (2013, p. 15) 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.” Similar goals or standards are promoted by the British Psychological Society [BPS] (2014) and other accrediting organizations. Collectively, these standards reflect a widely held understanding that an ability to source, read, understand and critically evaluate relevant research literature is a necessary precursor to evidence-based practice in psychology (American Psychological Association Presidential Task Force on Evidence Based Practice, 2006). The vast majority of this literature is based on quantitative research methods (Kidd, 2002; Rennie et al., 2002). It is also widely held that some of the most effective ways of teaching these skills involve engaging students regularly in all aspects of the research process, from the conception of meaningful research questions, through design, analysis, interpretation and reporting (Marek et al., 2004; Wagner et al., 2011; Earley, 2014; Stoloff et al., 2015). Hence, nearly all psychology departments provide multiple opportunities for undergraduate students to conduct original empirical research, either individually or in collaboration with other students or faculty (Kierniesky, 2005; Perlman and McCann, 2005).

Despite their importance, and their prominence throughout psychology curricula, research methods and (particularly) statistics are recognized areas of weakness for many students (Garfield and Ahlgren, 1988; Murtonen and Lehtinen, 2003; Garfield and Ben-Zvi, 2007; Murtonen et al., 2008). 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 which has been referred to as ‘selection skill’ (Ware and 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-min period. The scenarios reflected statistical concepts typically found in introductory statistics textbooks and widely used in behavioral science research. Despite most students having completed at least six research methods and statistics units¹, they overwhelmingly found the task difficult and performed poorly. On average, students managed to read 10.9 scenarios within the allocated time, and answered 25.3% of them correctly. An additional 15.7% of answers were coded as ‘partially correct.’ When Gardner and Hudson questioned the students

about how they made their decisions, several explanations for the poor performance emerged. These included students misinterpreting the research scenarios, knowing but being unable to name appropriate statistics, misidentifying the measurement levels (e.g., nominal, ordinal, continuous) of variables, and seizing on misleading keywords and data presentation formats.

When Allen et al. (2016) presented similar research scenarios to undergraduate psychology students, they also found the task of identifying appropriate statistical tests and procedures particularly challenging. Many were apologetic, and expressed embarrassment at being unable to successfully complete a task they felt they ought to be equipped to accomplish. When prompted to think about the process of selecting a statistical procedure (rather than actually identifying one), they continued to struggle. The processes they described tended to be haphazard and inefficient, and included looking for clues in the wording of scenarios, searching through textbooks, relying on memory or simply guessing. Of those who recognized that a systematic decision making process could be followed; none could identify every factor that would require consideration, and most also focused on irrelevant or peripheral aspects of the scenarios.

When students are asked to recognize (rather than recall) appropriate statistics, their performance appears similarly underwhelming. For example, Ware and Chastain (1989, p. 225) developed an eight-item multiple-choice selection skill test, which they and colleagues believed contained “problems that students should be able to solve after completing [an] introductory statistics course.” When they administered the test to students at the conclusion of such a course, the students answered fewer than 45% of the items correctly. Ware and Chastain (1989, p. 226) attributed this poor performance, at least partially, to a curriculum which taught statistical techniques “one at a time,” and did not emphasize the development of selection skills. A number of other researchers have also recognized that having relatively few opportunities to practice selection skills could account for the difficulties that students experience when placed in situations where they must work out *which statistic* to use (e.g., Quilici and Mayer, 1996, 2002; Lovett and Greenhouse, 2000; Yan and Lavigne, 2014).

Although not many 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 to place greater emphasis on when to use various statistics, and less on computational procedures, they observed a significant improvement on their multiple-choice selection skill test. In a more controlled context, Quilici and Mayer (2002) demonstrated that it is possible to train students to focus on the structural (e.g., the nature of the independent and dependent variables, and the relationship between them) rather than surface-level (e.g., topic) features of basic research scenarios, and that doing so improved students’ abilities to correctly categorize scenarios according to how they would be analyzed. After training, students were also better able to generate new scenarios that could be analyzed using the same statistical procedures as existing scenarios. More recently, similar findings were reported by Yan and Lavigne (2014), who observed that providing students with worked examples emphasizing the structural features of

¹In Australia, a ‘unit’ is a single subject, typically taken alongside two or three others over a semester. This term is analogous to ‘course’ in the United States.

simple research scenarios improved students' performance on subsequent categorization tasks, as well as their ability to identify the structural features defining each category.

Together, these findings suggest that selection skills are underpinned by 'structural awareness' (Quilici and Mayer, 2002), which reflects an ability to disregard the surface features of a research scenario, and focus on its structural features and the relations between them. Like the worked examples used by Yan and Lavigne (2014), graphic organizers, particularly decision trees and flow charts, provide a pedagogical tool for systematically focusing attention on these structural features and relations.

GRAPHIC ORGANIZERS

Graphic organizers are known to facilitate the process of selecting appropriate statistical tests and procedures for different types of research questions and data. They focus the user on each structural component of a research scenario, and illustrate their connectedness/differentiation with spatial positioning and lines (Nesbit and Adesope, 2006). The structured nature of graphic organizers can help users organize new information and integrate it with existing knowledge into schemata (Yin, 2012). The grouping of information lessens cognitive load, and thus more working memory can be applied to learning and problem solving (Yin, 2012). Furthermore, graphic organizers encourage both verbal and spatial encoding of new information, thus providing multiple pathways for its later recall (Katayama and Robinson, 2000). Meta-analyses support the efficacy of concept maps, a type of graphic organizer, for increasing student achievement (Horton et al., 1993), knowledge retention and transfer (Nesbit and Adesope, 2006), and learning (Moore and Readence, 1984).

A number of different types of graphic organizers have been created to help students select appropriate statistical analyses, including tip sheets which sort analyses by their defining characteristics (e.g., Twycross and Shields, 2004), and charts which link statistics to common research goals (e.g., Beitz, 1998). However, the organizers which have gained most traction follow decision tree logic, and are designed to guide the user from an initial question (or problem) to an answer or outcome, via a series of choice or decision points. In domains that involve complex rules, procedures, conditions, and multiple candidate solutions, the use of a decision tree can provide a highly organized approach to the process of decision-making. In the domain of statistics, decision-trees to guide statistical decision making have a long history (e.g., Mock, 1972; Fok et al., 1995) and are now commonly included in statistics textbooks (see, for e.g., Tabachnick and Fidell, 2013; Allen et al., 2014). Statistical decision trees differ from other types of graphic organizers in that they are hierarchical and start with a single node before branching off. By following the branches that refer to the key structural details of a research scenario, the user is led to a statistical analysis appropriate to their circumstances (Mertler and Vannatta, 2002). Theoretically, decision trees rest on the idea that knowledge must be organized or structured to be accessible from long-term memory (Schau and Mattern, 1997). Decision trees provide this structure by explicitly highlighting the interconnectedness (and

differentiation) between important statistical concepts (Schau and Mattern, 1997; Yin, 2012).

Empirically, there is work illustrating both the objective efficacy of statistical decision trees, as well as their subjective appeal. For example, Carlson et al. (2005; Protsman and Carlson, 2008) demonstrated that decision trees could facilitate significantly faster and more accurate (by a multiple of three) statistical decision-making, compared to more traditional methods of statistical test selection (e.g., by searching through a familiar textbook). The decision tree method was also significantly more popular amongst students than the textbook method (Carlson et al., 2005; Protsman and Carlson, 2008).

Despite their popularity, traditional statistical decision trees also 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 make traversing the tree easier are either spatially separated from the tree itself, or completely absent (Koch and Gobell, 1999; Blankenship and Dansereau, 2000). Second, when given to students without accompanying resources (e.g., a textbook) they do not provide sufficient detail to execute and interpret the statistics they help identify. Third, while the complexity and non-linearity 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 and Dansereau, 2000; Nesbit and Adesope, 2011).

To overcome these limitations, a number of researchers and educators have adapted the traditional decision tree model for digital media. These hypertext systems are typically comprised of a series of interconnected pages or nodes (Unz and Hesse, 1999). Space constraints associated with paper decision trees are removed, and links can be made to external resources that aid learning (Koch and Gobell, 1999). Map shock can be eliminated because the user is only shown a small section of the tree at any given time, reducing its complexity and ability to overwhelm (Blankenship and 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 'lost in hyperspace' (Otter and 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 in comparison to textual interfaces. In particular, when compared to textual interfaces, graphical map interfaces are associated with more effective (medium to large effect sizes) and efficient (small to medium effect sizes) performance (Chen and Rada, 1996).

Koch and Gobell (1999) adapted paper decision trees for delivery on the world-wide-web, and in doing so were able to also provide users with definitions, links to online resources, and information about how to enter and analyze data in commonly used statistical software. Like Carlson et al. (2005; 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 that provided by University of California, Los Angeles (UCLA)'s Institute for Digital Research and Education at http://www.ats.ucla.edu/stat/mult_pkg/whatstat/default.htm. This 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 format) 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 software applications currently available (e.g., Subramanian, 2014; Wacharamanotham et al., 2015) automatically select the statistical test for the user without explicitly guiding the user through the steps to make the decision, greatly reducing their pedagogic potential. STestMAP (Eng et al., 2011) is a visual tool that guides students through a systematic process to select a statistical test, but does not appear to be publicly available. Despite their potential benefits, hypertext decision trees and currently available software generally 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, 2011, cited in Yu et al., 2014, p. 2126). In the previous decade, the use of mobile learning technologies such as smart devices and mobile applications has increased rapidly, and amongst western higher education students their penetration is near ubiquitous (Stowell, 2011; Murphy et al., 2013; Dahlstrom and Bichsel, 2014; Chen et al., 2015). Their broad appeal is tied to many factors, including portability, enabling the user to access information and resources virtually anywhere and at any time (Jeng et al., 2010), and utility. Increasingly, students prefer to use their own smart devices for learning, and mobile learning applications have been identified as one of the technologies expected to have the biggest impact on education this decade (Martin et al., 2011; Johnson et al., 2012). In the context of teaching research methods and statistics, emerging research suggests that technology assisted examples delivered via mobile applications positively impact on student learning (Harnish et al., 2012).

STATHAND: A MOBILE APPLICATION TO SUPPORT STATISTICAL DECISION MAKING

In the previous sections of this paper, we have described how students find statistical test selection difficult, argued that decision trees can facilitate this decision making process, and

noted the rapid adoption of smart devices and mobile learning applications in the higher education sector. With these points in mind, we proposed StatHand to the Australian Government Office for Learning and Teaching in 2013. StatHand was described as a cross-platform mobile application that helps users quickly identify appropriate statistical tests and analytic procedures for a wide range of research questions, hypotheses and data types. The proposal, to develop, disseminate and evaluate StatHand, was funded.

The content of StatHand is being developed in two main phases. The first phase, which is now complete, is focused on helping users identify statistical tests and procedures appropriate to a wide range of circumstances. It is freely available in the iOS App Store, and can also be accessed as a fully mobile-compatible web application at <https://stathand.net>. The second phase, which is currently under development, guides the computation, interpretation and reporting of these tests and procedures.

The first phase of content is illustrated in **Figure 1**, on the iOS iPhone application. When StatHand 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: 'describe a sample,' 'compare samples,' 'analyze relationships or associations between variables,' 'examine the underlying structure of a measure,' and 'examine the reliability of a measuring instrument.' The statistics, tests and procedures under each of these objectives are listed in **Table 1**. Let's imagine that 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 prior to testing. The most appropriate option on Screen 1 is 'compare samples,' as we wish to compare the performance of the coffee drinkers with that of the water drinkers. After making our first selection, we are presented with a second choice, in which we need to identify the number of dependent variables in the study. A user uncertain about what is meant by 'dependent variable' can consult the brief annotation below the question, whereas more experienced users can simply make their selection. Here, we indicate that we have 'one' dependent variable (Screen 2), which is measured on an 'interval or ratio' scale (Screen 3). Next, we are promoted to consider the number and nature of our independent variable(s). As illustrated in Screens 4 and 6, each option can be expanded for context-specific definitions and examples by tapping on the relevant Information icons. Finally, we are asked to indicate whether or not 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.

At any point during the decision making process, a user can review their previous choices using the History tool, as illustrated in Screen 9 of **Figure 2**. This feature allows the user to retrace their steps, and draw stronger connections between their choices and the solutions they reach. Selecting any entry

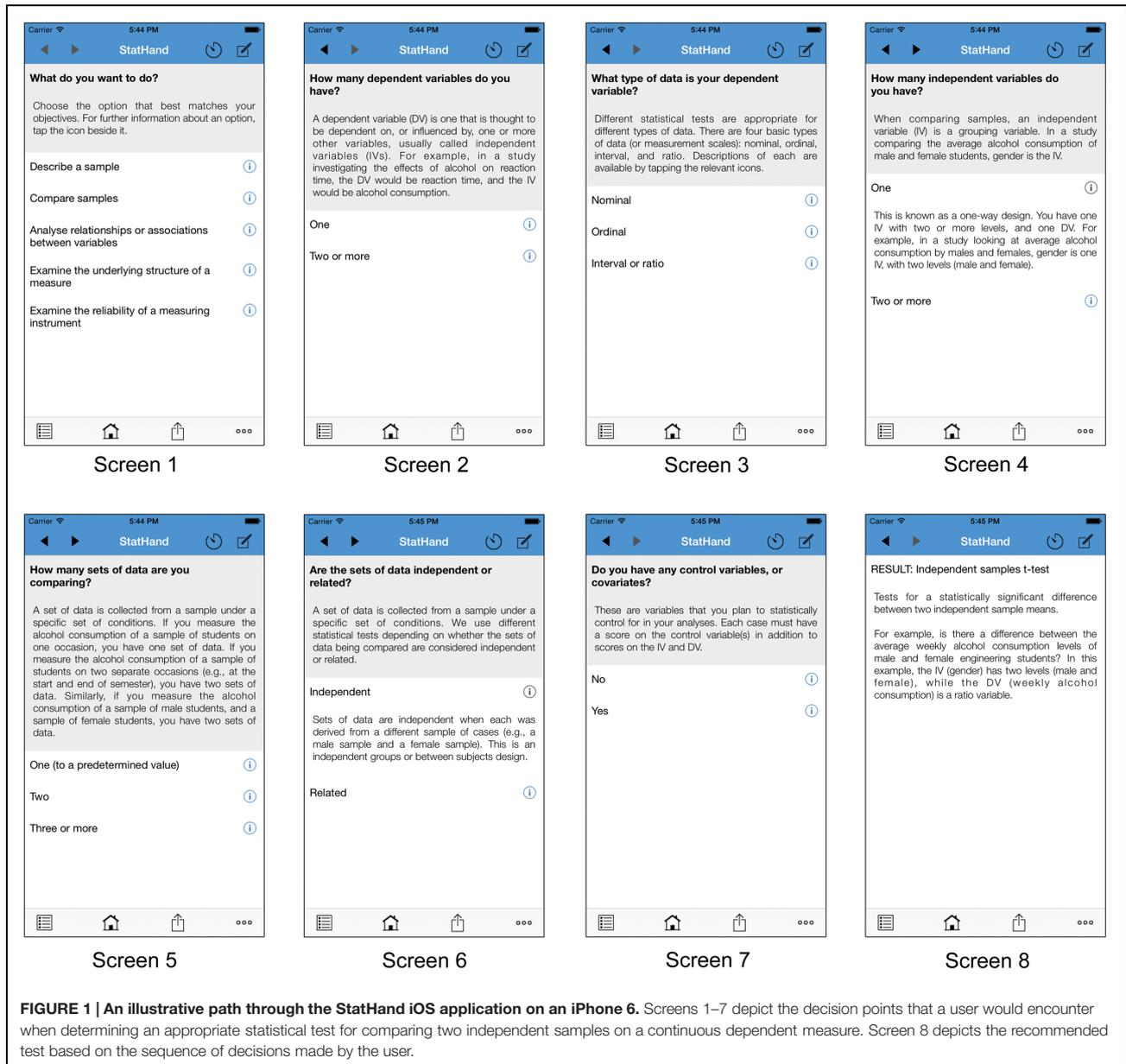


FIGURE 1 | An illustrative path through the StatHand iOS application on an iPhone 6. Screens 1–7 depict the decision points that a user would encounter when determining an appropriate statistical test for comparing two independent samples on a continuous dependent measure. Screen 8 depicts the recommended test based on the sequence of decisions made by the user.

in the History returns the user to the corresponding decision point. Users can also navigate through StatHand using the Back and Forward buttons, or jump directly to a statistic from the searchable Index (illustrated in Screen 10). Also illustrated in Screen 9, Figure 2 is the Notes tool, with which the user can pin their own annotations to specific pages within the application, or retrieve notes made on other pages. Finally, tapping on the Share icon in the toolbar at the bottom of the screen reveals options to print, email or save the annotated sequence of decisions leading to the current page (including the Notes associated with those decisions). It should be noted that these features work in comparable ways in the web version of StatHand at <https://stathand.net>, which has been designed for

compatibility with any device capable of running a modern web browser.

SUGGESTIONS FOR INTEGRATING STATHAND INTO THE RESEARCH METHODS CURRICULUM

As we've observed, many psychology students find the task of selecting appropriate statistics for different research questions, hypotheses and data types challenging (Gardner and Hudson, 1999; Allen et al., 2016). This selection skill (Ware and Chastain, 1989) appears underpinned by structural awareness (Quilici

TABLE 1 | The statistics, tests and procedures described in StatHand, grouped by research objective.

Objective	Statistics, tests and procedures described in StatHand
Describe a sample	Bar graph; category count; histogram; interquartile range; Mean; median; mode; pie chart; range; standard deviation; stem-and-leaf plot.
Compare samples	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).
Analyze relationships or associations between variables	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).
Examine the underlying structure of a measure	Confirmatory factor analysis; exploratory factor analysis; principal components analysis.
Examine the reliability of a measuring instrument	Cohen's kappa; Cronbach's alpha; intraclass correlation; Kuder-Richardson 20; Weighted kappa.

The objectives listed correspond with the five options presented to users on the StatHand home screen.

and Mayer, 2002, p. 326); an ability to disregard the surface features of research scenarios, and instead focus on their structural features and the relations between them. Traditional research methods and statistics courses underemphasize these skills, although research suggests that they can be trained (e.g., Quilici and Mayer, 2002; Yan and Lavigne, 2014). Decision trees provide a pedagogic tool for systematically focusing attention on the structural features of research scenarios, as well as the relations between them. StatHand reflects a new breed of interactive decision tree, ready for embedding in existing research methods and statistics curricula. It can be used to provide novel and engaging opportunities to practice selection skills and train students' structural awareness by systematically sensitizing them to the issues that require consideration before choosing between statistical techniques. Once the second phase content has been deployed, it can further be used as an aid to guide their computation, interpretation and reporting.

Research suggests that integrating technology generally (e.g., Tishkovskaya and Lancaster, 2012; Moreau, 2015), and mobile applications specifically (Harnish et al., 2012) into the research methods and statistics classroom can have pedagogic benefits. However, doing so is not without challenges. Potential barriers to successful integration include the limited confidence of teachers and students when working with new technologies, and differences in learning and teaching styles. Importantly, Lahiri and Moseley (2012, p. 11) cautioned that the use of smart devices as eLearning tools must be underpinned by pedagogical principles and an evidence base, otherwise the use of such

tools "might lead to frustration, inequity, shallow learning, and distraction from the main purpose of enhancing learning and making students competent professionals." Thus, in order to reduce students' statistics anxiety and enhance students' selection skills, teachers may wish to consider carefully how to effectively use smart devices as part of the learning process. Yu et al. (2014) stress that smart devices need to be used to extend the reach of teaching. Consequently, "shifting from e-learning to mobile learning implies that instructional designers need to adopt new ways of facilitating learning, not in one way, but using multiple pedagogical strategies, to help people learn whenever they need and wherever they are" (Yu et al., 2014, p. 2132).

StatHand was developed within the theoretical framework of the Unified Theory of Acceptance and Use of Technology (Venkatesh et al., 2003). This theory posits that performance expectancy, effort expectancy, social influence and facilitating conditions are direct determinants of the intention to use a particular technology, with intention and facilitating conditions predictors of actual use. Below we offer some suggestions for embedding StatHand in research methods and statistics courses.

Demonstrate StatHand at the Outset and Throughout the Course

StatHand is easily and freely accessible via the iOS App Store and online at <https://stathand.net>. Navigation through the application is intuitive (although brief instructions are available within the application), and largely self-contained, with definitions and examples of all key terms available at a simple tap of an icon. These features increase effort expectancy (defined in terms of ease of use, Venkatesh et al., 2003) Nevertheless, to maximize the application's perceived utility to students (part of performance expectancy), instructors should devote class time early in the semester to demonstrating how and when to use it. Revisiting StatHand each time a new analysis is introduced will help sensitize students to the similarities and differences between tests vis-à-vis their key structural characteristics (e.g., the key structural difference between the independent samples *t*-test and ANOVA is the number of levels of the independent variable). Such sensitivity is key to structural awareness, and the development of selection skills. Some instructors already use traditional (paper based) decision trees in efforts to achieve this aim. The benefits of transitioning to StatHand include the reduced potential for map-shock or 'glossing over key decision points,' the provision of an additional set of examples that students can refer to when seeking to master complex concepts, and the ability for students to save, print or email a record of their sequence of decisions (and annotations associated with those decisions) for later reference. Performance expectancy will increase as students succeed in selecting appropriate statistical techniques using StatHand.

Link StatHand to Existing Teaching Resources

StatHand can be easily incorporated into existing teaching activities and resources. For example, one of us (NL) created

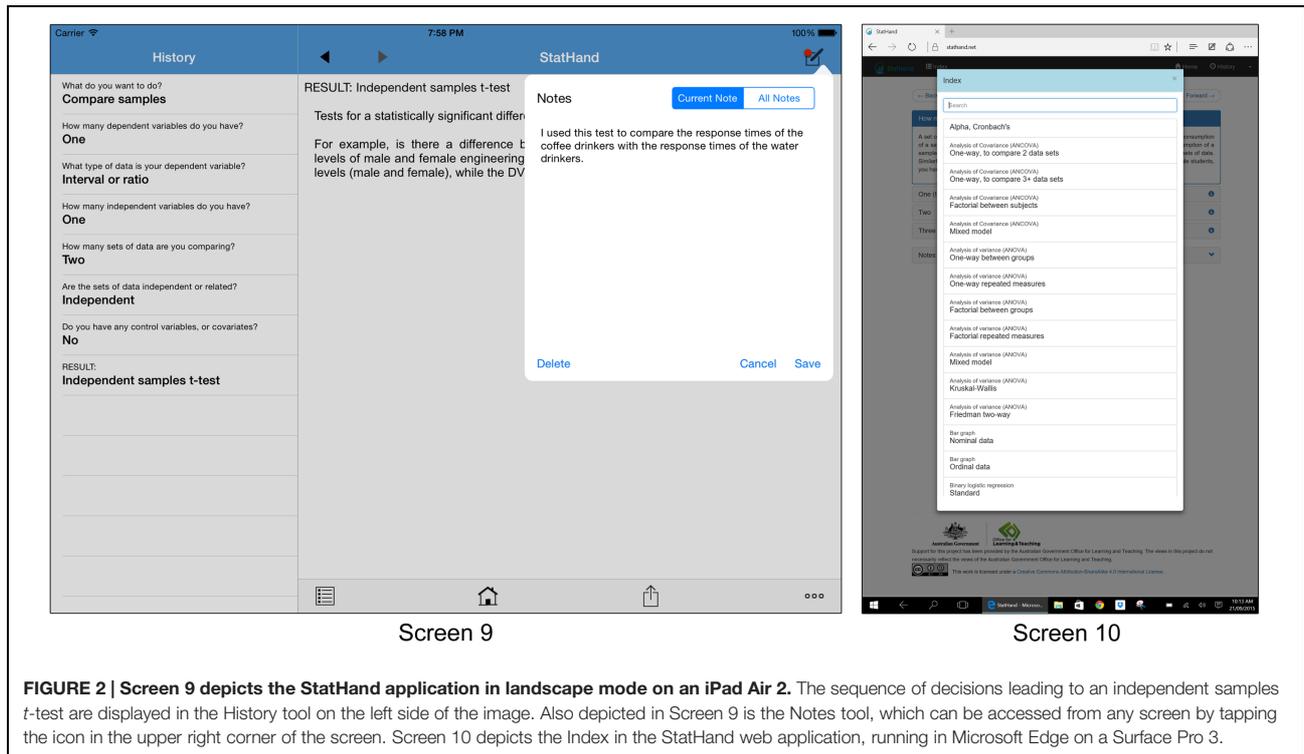


FIGURE 2 | Screen 9 depicts the StatHand application in landscape mode on an iPad Air 2. The sequence of decisions leading to an independent samples *t*-test are displayed in the History tool on the left side of the image. Also depicted in Screen 9 is the Notes tool, which can be accessed from any screen by tapping the icon in the upper right corner of the screen. Screen 10 depicts the Index in the StatHand web application, running in Microsoft Edge on a Surface Pro 3.

a YouTube screencast demonstrating the use of StatHand and embedded a link to the screencast (along with links to StatHand) in an existing worksheet demonstrating how to perform and interpret a specific statistical procedure. Another of us (PA) regularly uses it in tutorial activities and assessments, where students are presented with a research scenario and data set, and required to generate meaningful hypotheses. StatHand is then used to identify appropriate hypothesis tests, which are conducted and interpreted in the remainder of the class. The linking of StatHand to existing teaching resources combined with the annotated question feature of the StatHand app provide organization and technical infrastructure (facilitating conditions) to support adoption and use. The use of StatHand within existing forums such as discussion boards and social media sites facilitates social influence, particularly if used across multiple courses within the student’s degree.

Minimize Competition from other Sources

Competition from other sources of interaction when using technology in the classroom can impact on focus. To limit such distractions, students will need to be given clear advice about how to maximize the benefits that can be derived from using learning technologies. At a minimum, this may include recommending turning on ‘airplane’ mode on smart devices, which will prevent them from receiving notifications, and reduce students’ temptation to check emails, browse the web or use social networking applications.

Use StatHand Consistently and Repeatedly Throughout the Course (and other Related Courses)

When used effectively, StatHand can reinforce information provided by instructors, and offer practical experience in determining appropriate analyses for a variety of different research scenarios. When used consistently through statistics courses, and when statistical decision-making is explicitly assessed, selection skills can be generalized to other research-related courses. As a single application available free on a wide variety of platforms, StatHand can be readily incorporated across multiple courses in statistics and other research-focused courses throughout the psychology undergraduate degree. Over time, students will become increasingly familiar with StatHand, the promotion of its use by multiple instructors will enhance social influence, and both the intention to use, and actual use of StatHand. Its use will be second nature by the time they begin conducting individual (or small group) research projects in their final years of study.

FUTURE DIRECTIONS AND CONCLUSION

StatHand is a cross-platform application designed to aid the process of selecting statistical tests and procedures for a wide range of research scenarios. It is currently available in the iOS App Store and at <https://stathand.net>. StatHand can be easily integrated into existing teaching and learning activities, or used

as a base for the development of new activities focused on exploring the circumstances in which different statistics are appropriate.

Content for the second phase of StatHand is currently under development. When incorporated into the iOS and web applications, it will guide users through the computation, interpretation and reporting of each statistic that StatHand helps identify (see **Table 1**). It will also provide advice on testing assumptions and calculating and interpreting effect sizes where appropriate; offer links to additional reputable information about each technique; and highlight controversies and alternative approaches where applicable. Much of this material is being prepared as short videos, developed following evidence-based multimedia learning object design principles (e.g., Clark and Mayer, 2011).

We have also started integrating StatHand into our own research methods and statistics units. This is informing the development of a set of instructors' resources to complement StatHand. These resources will include 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 own teaching purposes. The package of activities will include multiple-choice quizzes that instructors can use to assess their students' abilities to identify appropriate statistical tests and procedures under a wide variety of circumstances. These will be provided in formats suitable for inclusion in worksheets and tests, as well as formats suitable for inclusion in PowerPoint presentations that either do or do not make use of common audience response technologies (e.g., Turning Point Keepad). When available, the StatHand instructors' resources will be provided freely, on request, to anyone who teaches research methods, statistics and related subjects at recognized higher education institutions.

Dissemination of StatHand is ongoing, and as its user base expands we are collecting usage data that will inform how the application may be optimized to facilitate learning and the decision making process. Additional research projects are experimentally investigating the instructional efficiency of StatHand relative to other common decision making aids (e.g.,

paper based decision trees and familiar textbooks). Further research will empirically investigate students' adoption and use of StatHand within the Unified Theory of Acceptance and Use of Technology framework (Venkatesh et al., 2003). Finally, we will soon begin investigating how instructors use StatHand to support the learning and teaching within their own courses. This multi-pronged evaluation approach has two ultimate aims. The first of these is to inform the ongoing development of StatHand. The second is to develop an evidence base and best-practice recommendations to guide its use.

To conclude, in this Technology Report we have provided an overview of StatHand, a free cross-platform mobile application designed to support students' statistical decision making. Developed with the support of the Australian Government Office for Learning and Teaching, StatHand guides users through a series of simple, annotated questions to help them identify a statistical test or procedure appropriate to their circumstances. In its next release, StatHand will also guide the computation, interpretation and reporting of the tests and procedures it helps users identify. We invite psychology research methods and statistics instructors to contact us about incorporating StatHand into their own classes.

AUTHOR CONTRIBUTIONS

PA led the development of the StatHand application, with support and contributions from LR, FB, NL, DVR, and AR. All authors contributed to the preparation of this manuscript.

ACKNOWLEDGMENTS

Support for this project has been provided by the Australian Government Office for Learning and Teaching (Grant Ref: ID13-2954). The views in this project do not necessarily reflect the views of the Australian Government Office for Learning and Teaching. The authors would also like to acknowledge Mortaza Rezae, Xavier Begue, and Ivan Dwiputera, who coded the StatHand IOS application and web application under the mentorship of Dr. Aneesh Krishna at Curtin University.

REFERENCES

- Allen, P., Bennett, K., and Heritage, B. (2014). *SPSS Statistics Version 22: A Practical Guide*. Melbourne, Australia: Cengage Learning.
- Allen, P. J., Dorozenko, K. P., and Roberts, L. D. (2016). Difficult decisions: a qualitative exploration of the statistical decision making process from the perspectives of psychology students and academics. *Front. Psychol.* 7:188. doi: 10.3389/fpsyg.2016.00188
- American Psychological Association Board of Educational Affairs Task Force on Psychology Major Competencies (2013). *APA Guidelines for the Undergraduate Psychology Major: Version 2.0*. Washington, DC: American Psychological Association Board of Educational Affairs Task Force on Psychology Major Competencies.
- American Psychological Association Presidential Task Force on Evidence Based Practice (2006). *Evidence-based practice in psychology*. *Am. Psychologist* 61, 271–285. doi: 10.1037/0003-066X.61.4.271
- Australian Psychology Accreditation Council [APAC] (2014). *Accreditation Standard for Programs of Study in Psychology Eligible for Approval Under the National Registration and Accreditation Scheme for the Health Professions: Second Consultation Draft*. Melbourne: APAC.
- Beitz, J. (1998). Helping students learn and apply statistical analysis: a metacognitive approach. *Nurse Educ.* 23, 49–51. doi: 10.1097/00006223-199801000-00016
- Blankenship, J., and Dansereau, D. (2000). The effect of animated node-link displays on information recall. *J. Exp. Educ.* 68, 293–308. doi: 10.1080/00220970009600640
- British Psychological Society [BPS] (2014). *Standards for the Accreditation of Undergraduate, Conversion and Integrated Masters Programmes in Psychology*. Leicester: BPS.

- Carlson, M., Protsman, L., and Tomaka, J. (2005). Graphic organizers can facilitate selection of statistical tests. *Part 1 - analysis of group differences. J. Phys. Ther. Educ.* 19, 57–66.
- Chen, B., Sellhamer, R., Bennett, L., and Bauer, S. (2015). *Students' Mobile Learning Practices in Higher Education: A Multi-Year Study. Educause Review*. Available at: <http://er.educause.edu/articles/2015/6/students-mobile-learning-practices-in-higher-education-a-multiyear-study>
- Chen, C., and Rada, R. (1996). Interacting with hypertext: a meta-analysis of experimental studies. *Hum. Comput. Interact.* 11, 125–156. doi: 10.1207/s15327051hci1102_2
- Clark, R. C., and Mayer, R. E. (2011). *E-Learning and the Science of Instruction*, 3rd Edn. San Francisco, CA: Pfeiffer.
- Cranney, J., and Dunn, D. S. (2011). Psychological literacy and the psychologically literate citizen: new frontiers for a global discipline,” *The Psychologically Literate Citizen: Foundations and Global Perspectives*, eds J. Cranney and D. S. Dunn (Oxford: Oxford University Press), 3–12.
- Dahlstrom, E., and Bichsel, J. (2014). *ECAR Study of Undergraduate Students and Information Technology, 2014*. Louisville, CO: Educause Center for Analysis and Research.
- Earley, M. A. (2014). A synthesis of the literature on research methods education. *Teach. High. Educ.* 19, 242–253. doi: 10.1080/13562517.2013.860105
- Eng, T. H., Hua, L. U., Kheng, F. K., Ying, L. C., and Yee Hui, R. C. (2011). Significance test selection via STTestMAP. *GSTF Business Rev.* 1, 155–160. doi: 10.5176_2010-4804_1.2.69
- Fok, L. Y., Angelidis, J. P., Ibrahim, N. A., and Fok, W. M. (1995). A decision tree approach to the interpretation of multivariate statistical techniques. *J. Educ. Business* 71, 110–117. doi: 10.1080/08832323.1995.10116770
- Gardner, P. L., and Hudson I. (1999). University students' ability to apply statistical procedures. *J. Stat. Educ.* 7.
- Garfield, J., and Ahlgren, A. (1988). Difficulties in learning basic concepts in probability and statistics: implications for research. *J. Res. Math. Educ.* 19, 44–63. doi: 10.2307/749110
- Garfield, J., and Ben-Zvi, D. (2007). How students learn statistics revisited: a current review of research on teaching and learning statistics. *Int. Stat. Rev.* 75, 372–396. doi: 10.1111/j.1751-5823.2007
- Harnish, D., Ling, C., and Shehab, R. (2012). Leveraging the use of mobile applications to increase knowledge retention in a classroom lecture. *Proc. Hum. Fact. Ergon. Soc. Annu. Meet.* 56, 610–614. doi: 10.1177/1071181312561127
- Horton, P., Mcconney, A., Gallo, M., Woods, A., Senn, G., and Hamelin, D. (1993). An investigation of the effectiveness of concept mapping as an instructional tool. *Sci. Educ.* 77, 95–111. doi: 10.1002/sci.3730770107
- Jeng, Y. L., Wu, T. T., Huang, Y. M., Tan, Q., and Yang, S. J. H. (2010). The add-on impact of mobile applications in learning strategies: a review study. *Educ. Technol. Soc.* 13, 3–11.
- Johnson, L., Adams, S., and Cummins, M. (2012). *Technology Outlook for Australian Tertiary Education 2012–2017: An NMC Horizon Report Regional Analysis*. Austin, TX: The New Media Consortium.
- Katayama, A., and Robinson, D. (2000). Getting students “partially” involved in note-taking using graphic organizers. *J. Exp. Educ.* 68, 119–133. doi: 10.1080/00220970009598498
- Kidd, S. A. (2002). The role of qualitative research in psychology journals. *Psychol. Methods* 7, 126–138. doi: 10.1037/1082-989X.7.1.126
- Kierniesky, N. C. (2005). Undergraduate research in small psychology departments: two decades later. *Teach. Psychol.* 32, 84–90. doi: 10.1207/s15328023top3202_1
- Koch, C., and Gobell, J. (1999). A hypertext-based tutorial with links to the Web for teaching statistics and research methods. *Behav. Res. Methods Instrum. Comput.* 31, 7–13. doi: 10.3758/BF03207686
- Kretser, H., Wong, R., Robertson, S., Pershyn, C., Huang, J., Sun, F., et al. (2015). Mobile decision-tree tool technology as a means to detect wildlife crimes and build enforcement networks. *Biol. Conserv.* 189, 33–38. doi: 10.1016/j.biocon.2014.08.018
- Lahiri, M., and Moseley, J. L. (2012). Is mobile learning the future of 21st century education? Educational considerations from various perspectives. *Educ. Technol.* 52, 3–13.
- Lovett, M. C., and Greenhouse, J. B. (2000). Applying cognitive theory to statistics instruction. *Am. Statist.* 54, 196–206. doi: 10.2307/2685590
- Marek, P., Christopher, A. N., and Walker, B. J. (2004). Learning by doing: research methods with a theme. *Teach. Psychol.* 31, 128–131. doi: 10.1207/s15328023top3102_6
- Martin, S., Diaz, G., Sancristobal, E., Gil, R., Castro, M., and Peire, J. (2011). New technology trends in education: seven years of forecasts and convergence. *Comput. Educ.* 57, 1893–1906. doi: 10.1016/j.compedu.2011.04.003
- McGovern, T. V., Corey, L., Cranney, J., Dixon, W. E. Jr., Holmes, J. D., Kuebli, J. E., et al. (2010). “Psychologically literate citizens,” *Undergraduate Education in Psychology: A Blueprint for the Future of the Discipline*, ed. D. F. Halpern (Washington, DC: American Psychological Association), 9–27.
- Mertler, C., and Vannatta, R. (2002). *Advanced and Multivariate Statistical Methods: Practical Application and Interpretation*, 2nd Edn. Glendale, CA: Pyczak.
- Mock, T. J. (1972). A decision tree approach to the methodological decision process. *Account. Rev.* 47, 826–829.
- Moore, D. W., and Readence, J. F. (1984). A quantitative and qualitative review of graphic organizer research. *J. Educ. Res.* 78, 11–17. doi: 10.1080/00220671.1984.10885564
- Moreau, D. (2015). When seeing is learning: dynamic and interactive visualizations to teach statistical concepts. *Front. Psychol.* 6:342. doi: 10.3389/fpsyg.2015.00342
- Murphy, A., Farley, H., and Koronios, A. (2013). “Understanding the use of smart mobile technologies for learning in higher education,” in *Proceedings of Ascilite 2013 Sydney on Electric Dreams*, eds H. Carter, M. Gosper, and J. Hedberg, (Sydney: Ascilite shaping the future of tertiary education), 18–22.
- Murtonen, M., and Lehtinen, E. (2003). Difficulties experienced by education and sociology students in quantitative methods courses. *Stud. High. Educ.* 28, 171–185. doi: 10.1080/0307507032000058064
- Murtonen, M., Olkinuora, E., Tynjala, P., and Lehtinen, E. (2008). “Do I need research skills in working life?”: University students' motivation and difficulties in quantitative methods courses. *High. Educ.* 56, 599–612. doi: 10.1007/s10734-008-9113-9
- Nesbit, J., and Adesope, O. (2006). Learning with concept and knowledge maps: a meta-analysis. *Rev. Educ. Res.* 76, 413–448. doi: 10.3102/00346543076003413
- Nesbit, J., and Adesope, O. (2011). Learning from animated concept maps with concurrent audio narration. *J. Exp. Educ.* 79, 209–230. doi: 10.1080/00220970903292918
- Otter, M., and Johnson, H. (2000). Lost in hyperspace: metrics and mental models. *Interact. Comput.* 13, 1–40. doi: 10.1016/S0953-5438(00)00030-8
- Perlman, B., and McCann, L. I. (1999). The most frequently listed courses in the undergraduate psychology curriculum. *Teach. Psychol.* 26, 177–182. doi: 10.1207/S15328023TOP260302
- Perlman, B., and McCann, L. I. (2005). Undergraduate research experiences in psychology: a national study of courses and curricula. *Teach. Psychol.* 32, 5–14. doi: 10.1207/s15328023top3201_2
- Protsman, L., and Carlson, M. (2008). Graphic organizers can facilitate selection of statistical tests: Part 2 – Correlation and regression analysis. *J. Phys. Ther. Educ.* 22, 36–41.
- Quilici, J. L., and Mayer, R. E. (1996). Role of examples in how students learn to categorize statistics word problems. *Educ. Psychol.* 88, 144–161. doi: 10.1037/0022-0663.88.1.144
- Quilici, J. L., and Mayer, R. E. (2002). Teaching students to recognize structural similarities between statistics word problems. *Appl. Cogn. Psychol.* 16, 325–342. doi: 10.1002/acp.796
- Rennie, D. L., Watson, K. D., and Monteiro, A. M. (2002). The rise of qualitative research in psychology. *Can. Psychol.* 43, 179–189. doi: 10.1037/h0086914
- Roberts, L. D., Heritage, B., and Gasson, N. (2015). The measurement of psychological literacy: a first approximation. *Front. Psychol.* 6:105. doi: 10.3389/fpsyg.2015.00105
- Saville, B. K. (2008). *A Guide to Teaching Research Methods in Psychology*. Malden, MA: Blackwell.
- Schau, C., and Mattern, N. (1997). Use of map techniques in teaching applied statistics courses. *Am. Statistic.* 51, 171–175. doi: 10.1080/00031305.1997.10473955
- Stoloff, M. L., Good, M. R., Smith, K. L., and Brewster, J. (2015). Characteristics of programs that maximize psychology major success. *Teach. Psychol.* 42, 99–108. doi: 10.1177/0098628315569877

- Stowell, J. R., (2011). "Emerging technologies to improve teaching and learning in a digital world," in *Best Practices for Technology-Enhanced Teaching and Learning: Connecting to Psychology and the Social Sciences*, eds D. S. Dunn, J. H. Wilson, J. E. Freeman, and J. R. Stowell (New York, NY: Oxford University Press), 299–316.
- Subramanian, K. (2014). "VisiStat: Visualization-driven, interactive statistical analysis," in *CHI'14 Extended Abstracts on Human Factors in Computing Systems* (New York, NY: Association for Computing Machinery), 987–992.
- Tabachnick, B. G., and Fidell, L. S. (2013). *Using Multivariate Statistics*, 6th Edn. Boston, MA: Pearson.
- Tishkovskaya, S., and Lancaster, G. A. (2012). Statistical education in the 21st century: a review of challenges, teaching innovations and strategies for reform. *J. Stat. Educ.* 20.
- Twycross, A., and Shields, L. (2004). Statistics made simple: part 4 choosing the right statistical test. *Paediatr. Nurs.* 16, 24–24. doi: 10.7748/paed2004.10.16.8.24.c944
- Unz, D., and Hesse, F. (1999). The use of hypertext for learning. *J. Educ. Comput. Res.* 20, 279–295. doi: 10.2190/FB08-5H94-3R3Y-KQLY
- Venkatesh, V., Morris, M. G., Davis, G. B., and Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. *MIS Quarter.* 27, 425–478.
- Wacharamanotham, C., Subramanian, K., Völkel, S. T., and Borchers, J. (2015). "Statsplorer: Guiding novices in statistical analysis," in *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (New York, NY: Association for Computing Machinery), 2693–2702. doi: 10.1145/2702123.2702347
- Wagner, C., Garner, M., and Kawulich, B. (2011). The state of the art of teaching research methods in the social sciences: Towards a pedagogical culture. *Stud. High. Educ.* 36, 75–88. doi: 10.1080/03075070903452594
- Ware, M. E., and Chastain, J. D. (1989). Computer-assisted statistical analysis: a teaching innovation? *Teach. Psychol.* 16, 222–227. doi: 10.1207/s15328023top1604_16
- Ware, M. E., and Chastain, J. D. (1991). Developing selection skills in introductory statistics. *Teach. Psychol.* 18, 219–222. doi: 10.1207/s15328023top1804_4
- Yan, J., and Lavigne, N. C. (2014). Promoting college students' problem understanding using schema-emphasizing worked examples. *J. Exp. Educ.* 82, 74–102. doi: 10.1080/00220973.2012.745466
- Yin, Y. (2012). Using tree diagrams as an assessment tool in statistics education. *Educ. Assess.* 17, 22–49. doi: 10.1080/10627197.2012.697850
- Yu, C., Lee, S. J., and Ewing, C. (2014). "Mobile learning: emerging trends, issues, and challenges in teaching and learning," in *Proceedings of E-Learn: World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education 2014*, ed. T. Bastiaens (Chesapeake, VA: Association for the Advancement of Computing in Education), 2126–2136

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2016 Allen, Roberts, Baughman, Loxton, Van Rooy, Rock and Finlay. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Appendix A: Author Contribution Statements



31 May 2016

To whom it may concern,

I, Peter Allen, was a major contributor to the conceptualisation and coordination of the research resulting in the following paper:

Roberts, L. D., & Allen, P. J. (2012). Student perspectives on the value of research participation. In S. McCarthy, K. L. Dickson, J. Cranney, A. Trapp, & V. Karandashev (Eds.), *Teaching psychology around the world* (vol 3, pp. 198-211). Newcastle, England: Cambridge Scholars.

I contributed to the drafting, writing and editing of the paper above, which is included in my PhD thesis. Commensurate with the extent of my contribution, I am the second author on this paper.

I also declare that the data reported in this paper were collected by Lynne Roberts and I in 2008, then first analysed and reported by a group of final year BPsych dissertation students at Curtin University in 2009, under the co-supervision of Lynne and I. These data were then re-analysed by Lynne and I in 2010, in the context of the research questions articulated in the abovementioned paper. This paper was then prepared for publication without reference to the original student dissertations.

Peter Allen:  _____

Date: 31 May 2016

I, Lynne Roberts, endorse Peter Allen's contribution to the abovementioned paper, as specified above.

Assoc. Prof. Lynne Roberts:  _____

Date: 31/5/2016



31 May 2016

To whom it may concern,

I, Peter Allen, was a major contributor to the conceptualisation and coordination of the research resulting in the following paper:

Roberts, L. D., & **Allen, P. J.** (2013). A brief measure of student perceptions of the educational value of research participation.
Australian Journal of Psychology, 65, 22-29. doi:10.1111/ajpy.12007

I contributed to the drafting, writing and editing of the paper above, which is included in my PhD thesis. Commensurate with the extent of my contribution, I am the second author on this paper.

Peter Allen:  _____

Date: 31 May 2016

I, Lynne Roberts, endorse Peter Allen's contribution to the abovementioned paper, as specified above.

Assoc. Prof. Lynne Roberts:  _____

Date: 31 May 2016



31 May 2016

To whom it may concern,

I, Peter Allen, was a major contributor to the conceptualisation and coordination of the research resulting in the following paper:

Allen, P. J., & Baughman, F. D. (2016). Active learning in research methods classes is associated with higher knowledge and confidence, though not evaluations or satisfaction. *Frontiers in Psychology, 7*, Article 279. doi:10.3389/fpsyg.2016.00279

I made a significant contribution to the drafting, writing and editing of the paper above, which is included in my PhD thesis. Commensurate with the extent of my contribution, I am the first author on this paper.

Peter Allen:  _____

Date: 31 May 2016

I, Frank Baughman, endorse Peter Allen's contribution to the abovementioned paper, as specified above.

Dr Frank Baughman:  _____

Date: 1 June 2016



31 May 2016

To whom it may concern,

I, Peter Allen, was a major contributor to the conceptualisation and coordination of the research resulting in the following paper:

Allen, P. J., Lourenco, A., & Roberts, L. D. (2016). Detecting duplication in students' research data: A method and illustration. *Ethics and Behavior*, 26, 300-311. doi: 10.1080/10508422.2015.1019070

I made a significant contribution to the drafting, writing and editing of the paper above, which is included in my PhD thesis. Commensurate with the extent of my contribution, I am the first author on this paper.

Peter Allen:  _____

Date: 31 May 2016

I, Amanda Lourenco, endorse Peter Allen's contribution to the abovementioned paper, as specified above.

Amanda Lourenco:  _____

Date: 01-06-2016

I, Lynne Roberts, endorse Peter Allen's contribution to the abovementioned paper, as specified above.

Assoc. Prof. Lynne Roberts:  _____

Date: 31 May 2016



31 May 2016

To whom it may concern,

I, Peter Allen, was a major contributor to the conceptualisation, coordination and supervision of the research resulting in the following paper:

Allen, P. J., & Roberts, L.D. (2016). The impact of academic sponsorship on Web survey dropout and item non-response. *First Monday, 21*(2). doi: 10.5210/fm.v21i2.6144

I made a significant contribution to the drafting, writing and editing of the paper above, which is included in my PhD thesis. Commensurate with the extent of my contribution, I am the first author on this paper.

I also declare that the data reported in this paper were collected by final year BPsych students at Curtin University in 2010 (Experiment 1) and 2011 (Experiment 2) under my supervision. Aspects of these data were then used by the BPsych students to complete their respective dissertations. These data were then pooled and re-analysed by Lynne Roberts and I, in the context of the research questions articulated in the aforementioned paper. The paper was then prepared without reference to the original student dissertations.

Peter Allen:  _____

Date: 31 May 2016

I, Lynne Roberts, endorse Peter Allen's contribution to the abovementioned paper, as specified above.

Assoc. Prof. Lynne Roberts:  _____

Date: 31 May 2016



31 May 2016

To whom it may concern,

I, Peter Allen, was a major contributor to the conceptualisation of the following paper:

Allen, P. J., & Roberts, L. D. (2010). The ethics of outsourcing online survey research. *International Journal of Technoethics*, 1(3), 35-48. doi:10.4018/jte.2010070104

I made a significant contribution to the drafting, writing and editing of the paper above, which is included in my PhD thesis. Commensurate with the extent of my contribution, I am the first author on this paper.

Peter Allen:  _____

Date: 31 May 2016

I, Lynne Roberts, endorse Peter Allen's contribution to the abovementioned paper, as specified above.

Assoc. Prof. Lynne Roberts:  _____

Date: 31 May 2016



31 May 2016

To whom it may concern,

I, Peter Allen, was a major contributor to the conceptualisation of the following paper:

Roberts, L. D., & **Allen, P. J.** (2015). Exploring ethical issues associated with using online surveys in educational research.
Educational Research and Evaluation, 21, 95-108. doi:10.1080/13803611.2015.1024421

I contributed to the drafting, writing and editing of the paper above, which is included in my PhD thesis. Commensurate with the extent of my contribution, I am the second author on this paper.

Peter Allen:  _____

Date: 31 May 2016

I, Lynne Roberts, endorse Peter Allen's contribution to the abovementioned paper, as specified above.

Assoc. Prof. Lynne Roberts:  _____

Date: 31 May 2016



31 May 2016

To whom it may concern,

I, Peter Allen, was a major contributor to the conceptualisation and coordination of the research resulting in the following paper:

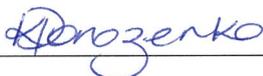
Allen, P. J., Dorozenko, K. P., & Roberts, L. D. (2016) Difficult decisions: A qualitative exploration of the statistical decision making process from the perspectives of psychology students and academics. *Frontiers in Psychology, 7*, Article 188. doi:10.3389/fpsyg.2016.00188

I made a significant contribution to the drafting, writing and editing of the paper above, which is included in my PhD thesis. Commensurate with the extent of my contribution, I am the first author on this paper.

Peter Allen:  _____

Date: 31 May 2016 _____

I, Kate Dorozenko, endorse Peter Allen's contribution to the abovementioned paper, as specified above.

Dr. Kate Dorozenko:  _____

Date: 1 June 2016 _____

I, Lynne Roberts, endorse Peter Allen's contribution to the abovementioned paper, as specified above.

Assoc. Prof. Lynne Roberts:  _____

Date: 31 May 2016 _____



31 May 2016

To whom it may concern,

I, Peter Allen, was a major contributor to the conceptualisation of the following paper:

Allen, P. J., Roberts, L. D., Baughman, F. D., Loxton, N. J., Van Rooy, D., Rock, A. J., & Finlay, J. (2016). Introducing StatHand: A cross-platform mobile application to support students' statistical decision making. *Frontiers in Psychology*, 7, Article 288. doi:10.3389/fpsyg.2016.00288

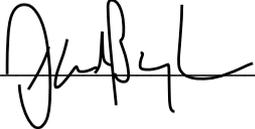
I made a significant contribution to the drafting, writing and editing of the paper above, which is included in my PhD thesis. Commensurate with the extent of my contribution, I am the first author on this paper.

Peter Allen:  _____ Date: 31 May 2016

I, Lynne Roberts, endorse Peter Allen's contribution to the abovementioned paper, as specified above.

Assoc. Prof. Lynne Roberts:  _____ Date: 31 May 2016

I, Frank Baughman, endorse Peter Allen's contribution to the abovementioned paper, as specified above.

Dr. Frank Baughman:  _____ Date: 1 June 2016

I, Natalie Loxton, endorse Peter Allen's contribution to the abovementioned paper, as specified above.

Dr. Natalie Loxton:  _____ Date: 1 June 2016

I, Dirk Van Rooy, endorse Peter Allen's contribution to the abovementioned paper, as specified above.

Dr. Dirk Van Rooy:  _____ Date: 1 June 2016

I, Adam Rock, endorse Peter Allen's contribution to the abovementioned paper, as specified above.

Dr. Adam Rock:  _____ Date: 7 June 2016

I, James Finlay, endorse Peter Allen's contribution to the abovementioned paper, as specified above.

James Finlay:  _____ Date: 8 June 2016

Appendix B: Copyright Permissions

Teaching Psychology around the World:
Volume 3

Edited by

Sherri McCarthy, K. Laurie Dickson,
Jacquelyn Cranney, Annie Trapp
and Victor Karandashev

**CAMBRIDGE
SCHOLARS**

P U B L I S H I N G

Teaching Psychology around the World: Volume 3,
Edited by Sherri McCarthy, K. Laurie Dickson, Jacquelyn Cranney,
Annie Trapp and Victor Karandashev

This book first published 2012

Cambridge Scholars Publishing

12 Back Chapman Street, Newcastle upon Tyne, NE6 2XX, UK

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the British Library

Copyright © 2012 by Sherri McCarthy, Jacquelyn Cranney, K. Laurie Dickson, Annie Trapp
and Victor Karandashev and contributors

All rights for this book reserved. No part of this book may be reproduced, stored in a retrieval system,
or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or
otherwise, without the prior permission of the copyright owner.

ISBN (10): 1-4438-3448-3, ISBN (13): 978-1-4438-3448-3

Subject: RE: Reprint Permission
Date: Friday, 5 August 2016 4:04:48 pm Australian Western Standard Time
From: Lynne Roberts
To: Peter Allen
Attachments: image002.png, image003.png

Dear Peter

You have my permission, as joint copyright owner, to reproduce the following publication in both your PhD thesis and the Curtin University institutional repository.

Roberts, L. D., & Allen, P. J. (2012). Student perspectives on the value of research participation. In S. McCarthy, K. L. Dickson, J. Cranney, A. Trapp, & V. Karandashev (Eds.), *Teaching psychology around the world* (vol 3, pp. 198-211). Newcastle, England: Cambridge Scholars.

Regards, Lynne Roberts

A/Prof Lynne Roberts

Mon-Tue: Director of Research in Higher Education

Wed-Fri: School of Psychology and Speech Pathology

Faculty of Health Sciences

2013 OLT National Teaching Fellow <http://www.dissertationsupervision.org/>

Curtin Academy Fellow

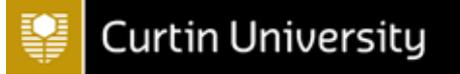
Curtin University

Tel | +61 8 9266 7183

Fax | +61 8 9266 2464

Email | Lynne.Roberts@curtin.edu.au

Web | <http://oasisapps.curtin.edu.au/staff/profile/view/Lynne.Roberts>



Curtin University is a trademark of Curtin University of Technology.
CRICOS Provider Code 00301J

From: Peter Allen
Sent: Friday, 5 August 2016 3:56 PM
To: Lynne Roberts <Lynne.Roberts@curtin.edu.au>
Subject: Reprint Permission

Dear Lynne,

I am seeking your permission, as joint copyright owner, to reproduce the following publication in (a) my PhD thesis, and (b) the Curtin University institutional repository at <https://espace.library.curtin.edu.au>.

Roberts, L. D., & Allen, P. J. (2012). Student perspectives on the value of research participation. In S. McCarthy, K. L. Dickson, J. Cranney, A. Trapp, & V. Karandashev (Eds.), *Teaching psychology around the world* (vol 3, pp. 198-211). Newcastle, England: Cambridge Scholars.

Regards, Peter Allen

Peter Allen

Lecturer | School of Psychology and Speech Pathology
Faculty of Health Sciences

Curtin University

Tel | +61 8 9266 2554

Fax | +61 8 9266 2464

Email | p.allen@curtin.edu.au

Web | <http://curtin.edu.au>



Curtin University is a trademark of Curtin University of Technology.
CRICOS Provider Code 00301J

NOTIFICATION OF ARCHIVING FOR INSTITUTIONAL REPOSITORIES – GRANTED

- This form notifies the APS of items planned for archiving in a registered institutional repository; refer to the 'Copyright permission request: Registration of institutional repository' form (available via the APS website) to register your repository with the APS.
- A separate form is required for each item archived.
- Please:
 - complete this form with the **work to be included in the institutional repository** and return the form to the Science and Education team at the APS (science@psychology.org.au)
 - return the form in **Microsoft Word format** (not PDF or scanned image)
- Refer to our website for more information (www.psychology.org.au/publications/copyright_request)

Institutional repository name	espace@Curtin
Name of author/s	Roberts, L. D., & Allen, P. J
Full citation of document	Roberts, L. D., & Allen, P. J. {2013}. A brief measure of student perceptions of the educational value of research participation. Australian Journal of Psychology, 65, 22-29. doi: 10.1111/ajpy.12007
Date of intended archiving	2016
Contact information	Amanda Bellenger, Manager: Strategy, Planning & Copyright Curtin University Library Library-Copyright@curtin.edu.au

(APS Use Only)

Payment requested by APS	<input type="checkbox"/> YES	If Yes, Amount: \$
	<input type="checkbox"/> NO	

Name	Sabine Hammond		
Position	Executive Manager – Science, Education and Membership		
Signature		Date	31 May 2016
Conditions:			
<ol style="list-style-type: none"> 1. Only an unenhanced copy of the postprint may be placed in the repository, without any links that would enhance use of the article 2. Acknowledgement of the first publication must be included, as follows: <p style="margin-left: 40px;">"This is an electronic version of an article published in [include the complete citation information for the final version of the article as published in the print edition of the journal]"</p> 			



Home

Create Account

Help



Title: Detecting Duplication in Students' Research Data: A Method and Illustration
Author: Peter J. Allen, Amanda Lourenco, Lynne D. Roberts
Publication: Ethics & Behavior
Publisher: Taylor & Francis
Date: May 18, 2016
Copyright © 2016 Routledge

[LOGIN](#)
If you're a copyright.com user, you can login to RightsLink using your copyright.com credentials. Already a **RightsLink user** or want to [learn more?](#)

Thesis/Dissertation Reuse Request

Taylor & Francis is pleased to offer reuses of its content for a thesis or dissertation free of charge contingent on resubmission of permission request if work is published.

[BACK](#)

[CLOSE WINDOW](#)

Copyright © 2016 [Copyright Clearance Center, Inc.](#) All Rights Reserved. [Privacy statement.](#) [Terms and Conditions.](#) Comments? We would like to hear from you. E-mail us at customercare@copyright.com

APPROVED

By Jan Travers at 1:37 pm, Jun 01, 2016

Submit by Email

Print Form

8/2/13

Request from Author for Reuse of IGI Materials

IGI Global ("IGI") recognizes that some of its authors would benefit professionally from the ability to reuse a portion or all of some manuscripts that the author wrote and submitted to IGI for publication. Prior to the use of IGI copyrighted materials in any fashion contemplated by the IGI Fair Use Guidelines for Authors, the author must submit this form, completed in its entirety, and secure from IGI the written permission to use such materials. Further, as a condition of IGI providing its consent to the reuse of IGI materials, the author agrees to furnish such additional information or documentation that IGI, in its sole discretion, may reasonably request in order to evaluate the request for permission and extent of use of such materials.

IGI will consider the special request of any author who:

- Completes, signs and returns this form agreeing to the terms; and
- Agrees that unless notified to the contrary, only the final, typeset pdf supplied by IGI Global is authorized to be posted (no pre-prints or author's own file.)

Title of article/chapter you are requesting: The ethics of outsourcing online survey research

Book Title and author/editor where this IGI material appears: International Journal of Technoethics
Vol 1, Issue 3, pp. 35-48

Purpose of request (where this material will appear):

- Posted on a secure university website for your students to access in support of a class. (Posted paper must carry the IGI Global copyright information as outlined above.)
- Posted in a university archive. The Website address is: http://espace.library.curtin.edu.au
- Posted on a personal Website: The Website address is: http://
- Republished in a book of which I am the editor/author. Book title of proposed book: _____

Publisher of proposed book: _____

- Other purpose (please explain):
Inclusion in my PhD by supplication thesis, which will then be deposited in the above institutional repository.

With your signature below, you agree to the terms stated in the IGI Global Fair Use Guidelines. This permission is granted only when IGI returns a copy of the signed form for your files and the manuscript pdf.

Your name: Peter Allen

Your signature: 

Organization: School of Psychology and Speech Pathology, Curtin University

Address: GPO BOX U1987, Perth, WA, 6845

Fax: 9266 2464

E-mail: p.allen@curtin.edu.au

For IGI Use

Request accepted by IGI Global: _____

Date: _____

J. Travers

Digitally signed by J. Travers
DN: cn=J. Travers, o=IGI Global,
ou=Director, Intellectual Property
and Contracts, email=jtravers@igi-
global.com, c=US
Date: 2016.06.01 13:37:43 -04'00'

Please complete and mail or fax this request form to:

Jan Travers • IGI Global, 701 E Chocolate Avenue • Hershey PA 17033 • Fax: 717/533-8661 • jtravers@igi-global.com



Home

Create Account

Help



Title: Exploring ethical issues associated with using online surveys in educational research

Author: Lynne D. Roberts, Peter J. Allen

Publication: Educational Research and Evaluation

Publisher: Taylor & Francis

Date: Feb 17, 2015

Copyright © 2015 Routledge

LOGIN

If you're a copyright.com user, you can login to RightsLink using your copyright.com credentials. Already **a RightsLink user** or want to [learn more?](#)

Thesis/Dissertation Reuse Request

Taylor & Francis is pleased to offer reuses of its content for a thesis or dissertation free of charge contingent on resubmission of permission request if work is published.

BACK

CLOSE WINDOW

Copyright © 2016 [Copyright Clearance Center, Inc.](#) All Rights Reserved. [Privacy statement.](#) [Terms and Conditions.](#) Comments? We would like to hear from you. E-mail us at customercare@copyright.com

Appendix C: List of Additional Peer-Reviewed Publications Since 2008

- Allen, P. J.** (2008) Rip, mix, burn ... sue ... ad infinitum: The effects of deterrence vs voluntary cooperation on non-commercial online copyright infringing behaviour. *First Monday*, 13(9). doi:10.5210/fm.v13i9.2073
- Allen, P. J.**, Shepherd, K., & Roberts, L. D., (2010). Peer-to-peer file-sharing: Psychological reactance and the theory of planned behaviour. *International Journal of Technoethics*, 1, 49-64. doi:10.4018/jte.2010100104
- Allen, P. J.**, Tonta, K. E., Haywood, S. B., Pereira, R. M., & Roberts, L. D. (in press). Predicting attendance at peer assisted study sessions: Supplementing the theory of planned behaviour with student role identity. *Active Learning in Higher Education*.
- Bánszki, F., Beilby, J., Quail, M., **Allen, P. J.**, Brundage, S. B., & Spitalnick, J. (in press). A clinical educator's experience using a virtual patient to teach communication and interpersonal skills. *Australasian Journal of Educational Technology*.
- Dzidic, P., Castell, E., Roberts, L. D., **Allen, P. J.**, & Quail, M. (2017). Reflections on the emergence and evolution of a scholarship of teaching and learning community of practice within a research-intensive higher education context. In J. McDonald & A. Carter-Steel (Eds.), *Communities of Practice - Facilitating Social Learning in Higher Education*. (pp. 219-239). Singapore: Springer Nature. doi:10.1007/978-981-10-2879-3_10
- Egan, S. J., Dick, M., & **Allen, P. J.** (2012). An experimental investigation of standard setting in clinical perfectionism. *Behaviour Change*, 29, 183-195. doi:10.1017/bec.2012.16
- Quail, M., Brundage, S. B., Spitalnick, J., **Allen, P. J.**, & Beilby, J. (2016). Student self-reported communication skills, knowledge and confidence across standardised patient, virtual and traditional clinical learning environments. *BMC Medical Education*, 16, Article 73. doi:10.1186/s12909-016-0577-5