



universität
wien

MASTERARBEIT / MASTER'S THESIS

Titel der Masterarbeit / Title of the Master's Thesis

„Student Samples in Psychological Research“

verfasst von / submitted by

Florian Stuefer, BSc

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of

Master of Science (MSc)

Wien, 2018 / Vienna 2018

Studienkennzahl lt. Studienblatt /
degree programme code as it appears on
the student record sheet:

A 066 840

Studienrichtung lt. Studienblatt /
degree programme as it appears on
the student record sheet:

Masterstudium Psychologie

Betreut von / Supervisor:

Univ.-Prof. Dr. Erich Kirchler

Table of Contents

Introduction	4
Differences between students and non-students	7
Method.....	8
Results	14
Categories	15
Discussion	17
Limitations	19
References.....	21
List of Figures	25
List of Tables	25
Appendix	26
Appendix A.....	26
Appendix B.....	28

Student Samples in Psychological Research

University and college students seem to be the sample of choice in many psychological studies. It is well documented that a big proportion of psychological research relies on university students for its experimental research (Landrum & Chastain, 1995). This is the case because students are easily available at universities where research is being conducted, questionnaire studies are sometimes done during university courses, students can be offered bonus points for various tests and courses to increase their grades or are even required to finish a course. Also, monetary incentives to participate in research are often too low to lure the employed population into the laboratory. Lastly, sometimes students are simply interested in research so they participate in studies. Just as the white laboratory rat is used in medical research which will be applied to humans, students are surrogating for consumers, business persons and many others in the psychological laboratory. Students are mostly used as research subjects because they are readily available and not because they are the population of interest. The problems which arise with this kind of research are, that students are rarely the population of interest, instead they are roleplaying or surrogating the actual population which is being investigated. A representative sample should be the consequence of random selection of subjects from the target population and random assignments of those subjects to the experimental groups. But because of logistic and financial limitations this approach is often impractical, so sampling is more often a convenience sample than the actual target population (Cook & Campbell, 1979). This approach might give rise to problems with reliability and external validity. External validity means the generalization of research findings from a sample to a larger population (Lucas, 2003). Reliability is defined as the extent to which repetition of a study would result in the same data and conclusions (Goode, 1952).

Peterson (2001) reported that 86% of the research subjects in empirical studies appearing in the Volume 26 of *Journal of Consumer Research* published in 2000, were students. Also 75% of the research subjects in *Journal of Consumer Research* and *Journal of Marketing Research* articles in the years 1995-1999 were college students (Simonson, Carmon, Dhar, Drolet & Nowlis, 2001). Espinosa and Ortinau (2016) analyzed 1090 Articles in the years 2009-2011 published in the *Journal of Consumer Research*, *Journal of Business Research*, *Journal of the Academy of Marketing Sciences* and the *Journal of Marketing*. The usage of student samples in this study ranged from 46% in the *Journal of Business Research*

to 96% in the *Journal of Consumer Research*. It is even argued that the excessive use of university students has led to a bias in psychological literature (Sears, 1986). This data which is built on university student data is often referred to as the “narrow database” (Sears, 1986). Many studies using convenience samples of students ignore the sample usage in the discussion or conclusion section, addressing to them as “participants”, “individuals” or “consumers” (Peterson & Merunka, 2014).

It is argued that university students are appropriate research subjects when the research is focused on basic psychological processes or theory testing linked to human behavior which is independent of sample characteristics (Lucas, 2003; Kardes, 1996). Mook (1983) argues that “representativeness of sample is of vital importance for certain purposes, such as survey research. For other purposes, it is a trivial issue” (p. 384). “Other purposes” means research that tries to draw conclusions about theory of human behavior, but not about a specific population. According to this, university students or other samples are appropriate as research subjects for basic research and theory testing (Mook, 1983). Students may also be used to replicate prior research conducted with student samples. Researchers which use convenience samples of university students usually use a single subject pool from one department or university, this might give rise to problems when replicating studies which used student samples with another, different student sample (Peterson & Merunka, 2014). University student subjects might enhance research validity because of their apparent homogeneity as students tend to be homogenous on dimensions such as age and education, which tends to influence attitudes (Peterson & Merunka, 2014). Such homogeneity is desirable as it decreases variability in measurements and increases the likelihood of rejecting a null hypothesis of no difference (Lynch, 1982). Also, homogenous samples increase the probability of identifying theory violations when a theory is false (Lucas, 2003). Arguments for or against student samples as research subjects focus on whether the results obtained from a sample are generalizable to non-student populations (Peterson & Merunka, 2014). An argument for student samples is that students can substitute another group properly when they possess similar experience and knowledge (Shuptrine, 1975). For example, older student samples responded the same to advertisements as other consumers, but younger student samples did not (Jones & Sonner, 2001). Between 1980 and 1987 the number of younger college students, 18 to 24 years, remained stable whereas the number of older college students increased according to American Demographics (as cited in Jones and Sonner, 2001). According to Jones and Sonner (2001) this shift of young students to older students happened because universities have expanded their class schedules, by including night classes

or weekend classes. This expanded the possibility to attend university courses for people who are employed. Due to this, students are no longer age-homogenous, consequently students should reflect the response behavior of the general public more correctly (Fuchs & Sarstedt, 2009). Also, in other, more specialized cases, the need for student samples that are relatively close to the actual sample which is being investigated arises. Students can represent non-students in specific cases, an example would be advanced business students as surrogates for executives or future leaders. Those are usually highly paid, with little motivation and time to participate in psychological research (Briggs, Balthazard & Dennis, 1996). For example, Remus (1996) compared the decision making of undergraduate and graduate business students to that of actual managers. This experiment showed that undergraduate students made significantly more costly decisions, used less effective heuristics and showed more erratic behavior, whereas the results from graduate students could be compared to real managers.

There are only few studies reviewing if student samples can be used as substitutes for non-students. Those studies are usually focused on specific areas like logistics research (e.g. Thomas, 2011), behavioral research (e.g. Ashton & Kramer, 1980; Remus 1996) or marketing research (e.g. Fuchs & Sarstedt, 2009). Fewer studies discuss the problem which could arise with student sampling in general (e.g. Hooghe, Stolle, Mahéo & Vissers, 2010; Newman, Joseph & Feitosa 2015). In various cases, the results obtained with standard convenience samples of undergraduate students couldn't be generalized toward other population groups (Hooghe et al., 2010). But still it is important to point out that the use of student samples should not be rejected across the board, as previous studies have shown that students can serve as appropriate surrogates, depending on the population being studied (Fuchs & Sarstedt, 2009). Those inconclusive results show that more research on the external validity of student data is needed. There is a gap of reviews comparing if students produce the same data under the same circumstances as non-students in psychological research. To fill this gap this review used studies which drew direct comparisons between students and non-students in the same situation, to determine if student samples produce the same data as a non-student sample. There are indices that students produce the same data as non-students if they are similar in age, experience, education or knowledge to the sample which is being investigated (Briggs et al., 1996; Shuptrine, 1975; Jones & Sonner, 2001). Therefore, this review also collected information about the type (e.g. questionnaire), the student and non-student sample (e.g. undergraduate business student and manager) and categories (e.g. consumer behavior). Those additional variables were used to find if there are systematic

variations between studies where students produced the same data as non-student, or not. For example, is student data more or less similar to non-student data in questionnaire studies or in research which is focused on consumer behavior or is a graduate student a better research subject than an undergraduate student.

Differences between students and non-students

If results obtained with student samples can't always be generalized towards other populations, university students need to be different in some variables compared to the general population.

Students tend to have weak self-definitions because of their limited life experience. Self-definition deals with a person's values, preferences, abilities and emotions which together form self-identity. These characteristics continue to develop as one matures (Ashraf & Merunka, 2017).

Compared with older adults, university students are likely to have less crystallized attitudes. Late adolescents and young adults have less crystallized social and political attitudes than older people do, also young people change attitudes more than older persons in response to political events (Sears, 1986). These unstable attitudes lead to a higher proneness to external influences, a weaker self-perception and attitude behavior inconsistency (Ashraf & Merunka, 2017).

Students have more unstable peer group relationships, which are not comparable to the long term social support and influence for judgements and attitudes of a more mature person. College students are less thoroughly tied to stable primary groups than young adults because they are more likely to leave those groups which they formed in their earlier life. They have not yet become fully embedded in the group relationships of their adulthood, such as in marriage, the workplace or the neighborhood (Sears, 1986).

Besides being confronted with academic evaluations, university students also have above average cognitive skills (Sears, 1986). Students are often confronted with all kinds of evaluations in their academic career. When a student is in a typical laboratory setting they may try to find the "right" answer and question the purpose of the study. This kind of thinking is probably unrepresentative for a setting in the real world like a consumer decision in the marketplace (Ashraf & Merunka, 2017).

Method

Systematic Quantitative Approach

This review is based on the approach of systematic literature reviews as described by Pickering and Byrne (2013) and the Preferred Reporting Items for Systematic Reviews and Meta-Analyses Statement (PRISMA) (Moher, Liberati, Tetzlaff & Altman, 2009). Pickering and Byrne (2013) argue that most scientific literature reviews are rather narrative, therefore researchers producing high quality literature reviews already need to be good writers and have some expertise in the field which they are reviewing. This method allows researchers or students new to a field, to systematically analyze existing academic literature to produce a structured quantitative summary of the topic. The difference between the traditional literature review and the systematic review, is what Pickering and Byrne (2013) have named *systematic quantitative approach*. This type of review is systematic because the methods used to survey the literature, then selecting the papers to include, are explicit and reproducible. This means similar results should be obtained if the procedure is repeated as described in the method section. This approach can be used as a different technique compared to the usual narrative review. Narrative reviews are very subjective, as the researcher can choose themselves which studies to include or not, which makes the review prone to several biases (Petticrew, 2001). The PRISMA flow diagram is a graphic illustration of how studies were selected (see Figure 1). This approach as described by Pickering and Byrne (2013) is following 15 consecutive steps describing the process of collecting and analyzing of the data (see Figure 2). The first six steps will be explained in detail as they belong to the methods section. The steps seven to 15 won't be explained further as they provide advice on how to write a systematic review, relevant an author.



PRISMA 2009 Flow Diagram

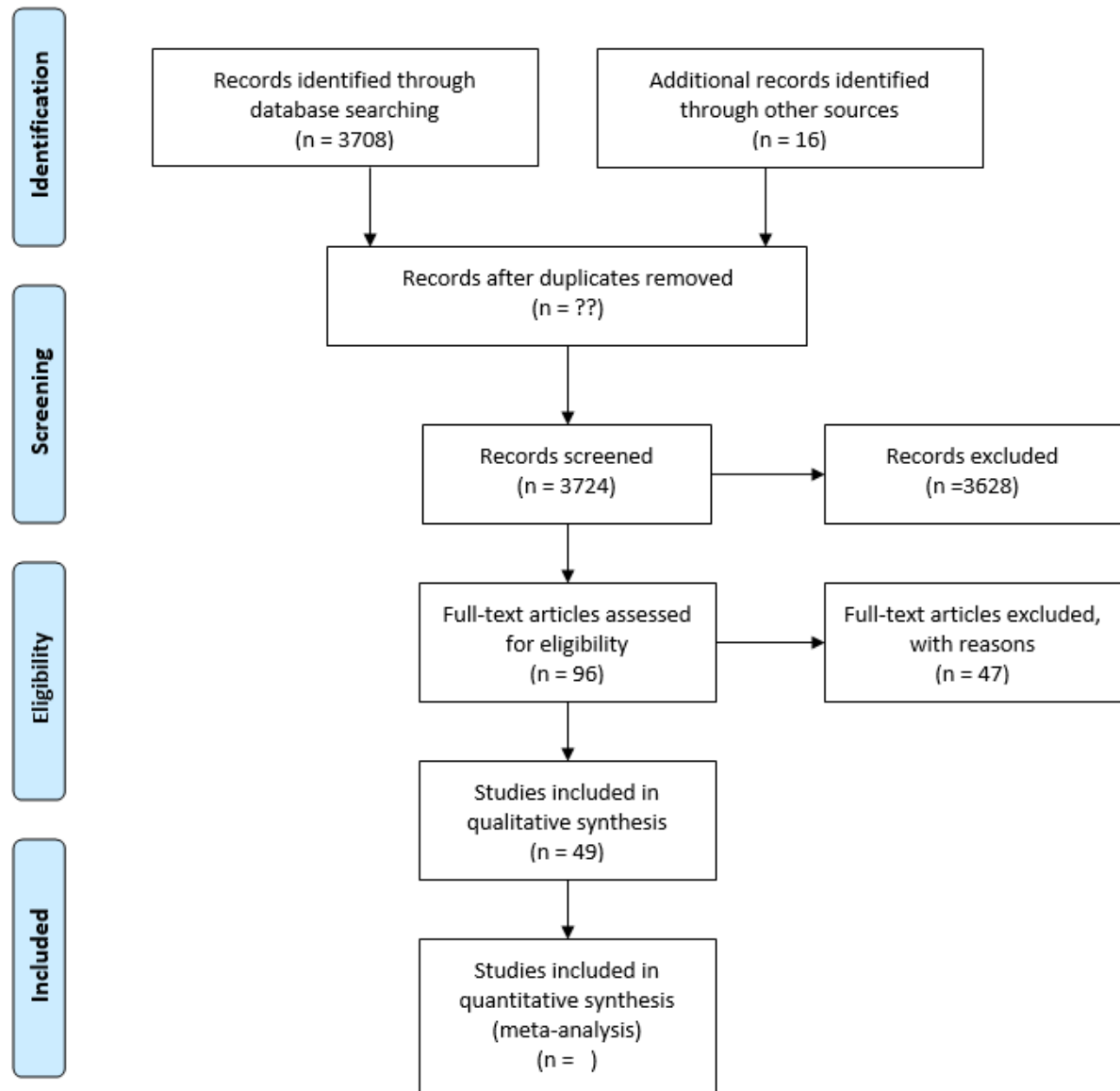


Figure 1. Prisma Flow Diagram. Adapted from “Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement,” by Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JPA, et al., 2009, *PLOS Med* 6(7), p. 4. Copyright 2009 by Liberati et al..

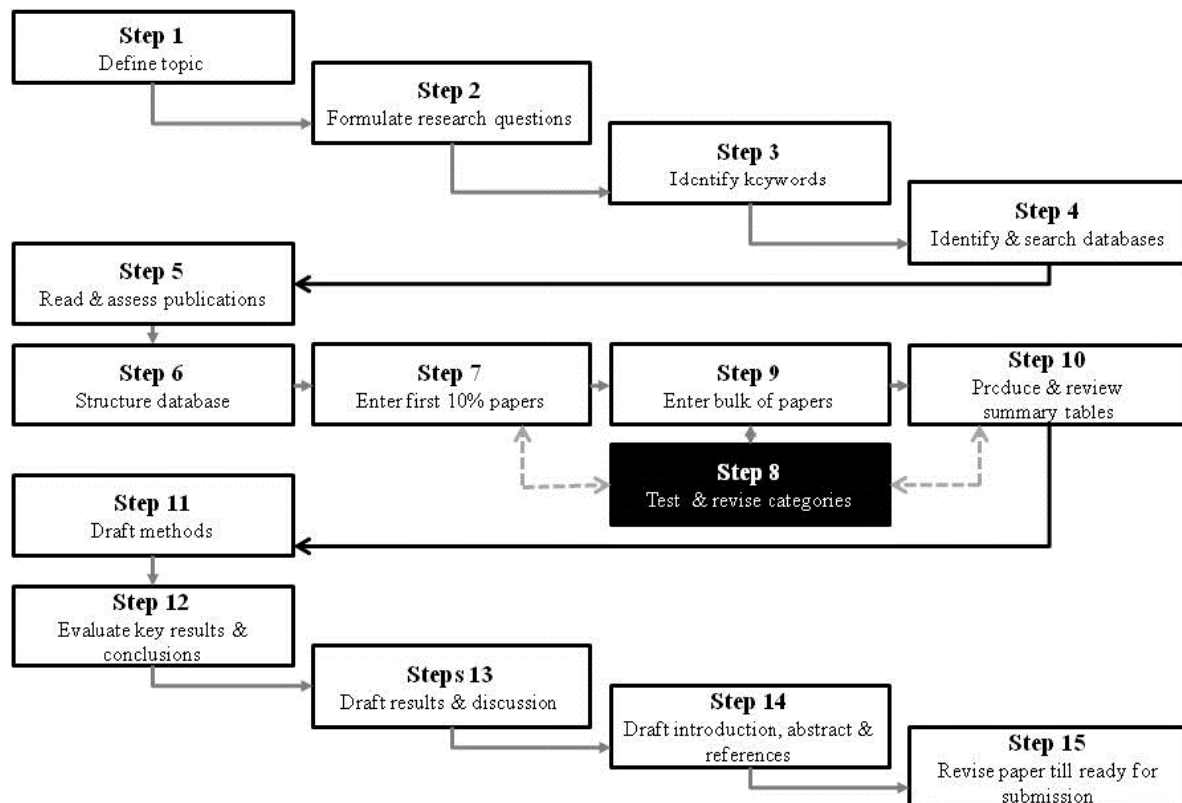


Figure 2. Fifteen stages in undertaking systematic quantitative literature reviews. Reprinted from “The benefits of publishing systematic quantitative literature reviews for PhD candidates and other early-career researchers,” by Pickering, C.M. and Byrne, J., 2013, *Higher Education Research and Development*, p. 10. Copyright 2013 by Taylor & Francis Group

Step 1: Define Topic. The first step is to define a specific topic within the field of research (Pickering & Byrne, 2013). As explained above this review tries to demonstrate if data from student samples differs from non-student sample data. Only research papers which provide comparisons between students and non-students will be included in the analysis, this does not include other literature reviews.

Step 2: Formulate research questions. It is important to identify what types of questions should be addressed by the literature review. This includes the types of subjects which were examined, which variables are of interest and the different disciplines assessing the topic (Pickering & Byrne, 2013). The first question is if students produce different data than non-student populations in psychological research. The second question is if there are systematic variations under which circumstances student samples produce the same, or different data than a non-student sample. Therefore, this review also includes information

about the type (e.g. questionnaire), the student and non-student sample (e.g. undergraduate business student and manager) and categories (e.g. consumer behavior).

Step 3: Identify Keywords. Once the topic and research questions have been defined key words need to be selected which will be used in the next step to search electronic databases for relevant papers (Pickering & Byrne, 2013). This literature search is based upon four articles by Ashton and Kramer (1980), Corrigan, Depositario, Nayga, Wu and Laude (2009), Liyanarachchi & Milne (2005) and Remus (1996). Based on those four articles a search for similar articles was conducted on google scholar with the “related articles”-function. Based on this search 12 articles discussing the issue of student samples in research were found. The keywords from those 16 research papers, 12 from google scholar plus four starting articles, were extracted and put in a frequency table. The most frequent keywords were “*student surrogates*” (five times), “*student samples*” (five times) and “*student*” (three times) which was removed because the word “*student*” was already included twice in the other keywords.

Step 4: Identify & search databases. Next appropriate databases need to be selected and searched (Pickering & Byrne, 2013). After a first test search with the keywords “student AND sample\$” finding 71987 results from Web of Science only, another approach was needed. Due to the high number of results but only very few articles which compare student data with non-student data this wasn’t an economic approach. This happened because as earlier mentioned, a lot of studies use student samples, but only a few investigate if students produce the same data as non-students. A lot of articles which investigate the problem if student samples are appropriate replacements for non-students, use the words “student” and “sample” in the title of the article. This led to the search strategy “student AND sample\$” in *Title* and “student AND surrogate\$” in the default search which is usually addressed as *all Fields* or *Topic*. The final search was conducted in the databases Web of Science (webofknowledge.com), Scopus (www.scopus.com) and APA PsycNET (psycnet.apa.org). These databases were selected because they search for quality psychological literature of all psychological disciplines with the possibility to search only in *Title*, as explained above. For example, PubMed is specialized on medical topics, therefore mostly clinical psychology. Other databases were excluded because they needed a paid subscription, such as PsycINFO. There is a lack of literature overall, therefore literature ranging from 1950 to 2017 was included. The results are displayed in the 2009 version of the PRISMA Flow Diagram (see Figure 1). Overall 3708 records were identified through database searching, of which 35

articles where found through “student AND surrogate\$” and 3673 through “student AND sample\$”, with the 16 starting articles this adds up to 3724 articles. Due to many thousand studies which were found there is no count of how many articles were left after removing duplicates. As there is also no value to this information this part of the flow diagram is left empty.

Step 5: Read & assess publications. This step involves reading and assessing each publication and deciding whether it should be included or not (Pickering & Byrne, 2013). A first screening for articles comparing student data with non-student data was done which excluded 3628 articles and left 96 articles. Those 96 articles still included reviews which won't be used in the final analysis as they don't compare student with non-student data themselves. After screening for original articles which provide information comparing student and non-student data the final number of studies included in this systematic literature review is 49, of which two studies provided two additional comparisons between students and non-students. Each study ($N=49$) had one data set, except for two studies, which each had one additional data set, so overall 51 data sets were analyzed. The first study with two data sets compared graduate and undergraduate students with a non-student sample. The second one compared students which received negotiation training, students without negotiation training and professional negotiators.

Step 6: Structure database. Based on the remaining publications a database needs to be structured (Pickering & Byrne, 2013). All data was put in an excel sheet in which the following variables were collected.

Type. To investigate if there are differences between types of data collection, the data was divided between the methods aggregate data, experiments and questionnaires. This structure was adapted from Muehlbacher and Kirchler (2016).

Samples. As there were indicators that not all student samples produce the same results (Briggs & et al., 1996; Shuptrine, 1975), data was collected on what kind of students were used. Examples given are graduate students, undergraduate students, business students or graduate business students.

Categories. To investigate if there are differences between different fields of psychological research the data was categorized. If at least two studies belonged to the same category, which couldn't be classified in another category, a new category was created. Following are the categories which were used.

- *Professional decision making*: Compares decision-making processes like investment decisions of professionals as managers or accountants with students in the same situation. Examples given are payroll decisions, selection decisions for applicants of managerial positions, negotiation behavior, salesman roleplaying, military interventions and implicit leadership theories.
- *Clinical/health*: This category includes studies which fall into clinical psychology, psychotherapy and health. Differences in scales for schizophrenia, dissociation, depression, fertility awareness and electrodermal indices to fear reaction. Comparisons regarding the effectiveness of psychotherapy and its underlying factor structure, and assessments of the suicide related knowledge of health professionals and students.
- *Consumer behavior*: Studies about a consumer's reaction to marketing and products. Examples given are differences in susceptibility to advertisements, willingness to buy in experimental auctions and attitudes towards online retailing services.
- *Social behavior*: Comparison of pro social behavior or other-regarding behavior in experiments. Examples given are trust experiments and if participants which participate in paid experiments behave differently than students which participate voluntarily in experiments.
- *Personality/attitudes*: General variables in personality such as the Big Five or differences in happiness, religiosity or motivation. Also attitudes of employees towards their organization and attitudes of students towards their university.
- *Ratings*: Studies in which data was rated or ranked by students and non-students. Includes a ranking of university mental health needs, such as interventions for problems with parents or depression. Another study ranked age perceptions of jobs, such as is a job position usually held by a younger, middle aged or older person.
- *Hospitality*: Studies about restaurants, hotels, tourism and how participants interact with the service staff. This category includes studies about differences in complaint behavior, as do customers complain to employees or managers after service failure and overall satisfaction with different hospitality service providers.
- *Criminology*: Studies about criminal psychology. Examples given are differences in training the ability to detect the right person from line ups after changing clothes or if there are population differences of the risk of stalking victimization.

- *Taxes*: Experiments about tax compliance of students and non-students in laboratory settings.

Results

A data set was categorized as positive, if the analyses in the original study of which the data was used, did not find significant differences between the student and the non-student sample, or the researchers indicated in other ways that the student sample does produce the same data as the non-student sample. A dataset was categorized as negative if there were significant differences between the student and the non-student sample data. For a better understanding how a study was rated as positive or negative, a few examples will be presented.

This experiment by Barr and Hitt (1986) from the category *Professional decision making* was rated as negative because “significant differences were noted between the two samples in the number and nature of factors used in making selection and salary decisions” (p. 599).

An example from the category *Consumer behavior* about the willingness to pay (WTP) (Depositario, Nayga, Wu & Laude, 2009) which was rated positive as their “results suggest that there is no significant difference between the WTP bids of the two groups” (p. 122).

A questionnaire study in the category *Ratings* about which university health needs are the most important by Henggeler, Sallis and Cooper (1980). This study was rated as negative because there was no agreement between the two samples because “the students rated problems of substance abuse as most serious, whereas the professionals rated problems of an academic and neurotic nature as most serious” (p. 217).

This questionnaire by Park and Lessig (1977) which compared housewives and students was rated negative because “the results reveal significant differences between housewives and students in terms of the influence which the three types of reference groups have upon brand selection” (p. 102).

And lastly, one study which generated two different data sets by Remus (1996) of which the undergraduate student data was rated as negative whereas the graduate student data was rated as positive. Their study showed that “there were no significant differences between

the managers and graduate business students. The undergraduate students, however, made more costly decisions, used less effective decision heuristics, and were more erratic than the managers and graduate students” (p. 93).

Overall 20 data sets indicated that student data is positive, and 31 data sets indicated that student data is negative.

Categories

The biggest category was *Professional decision making* which included 17 data sets. The smallest categories were social behavior, criminology and taxes with two data sets each. The methods used were 23 questionnaires, 24 experiments and four times aggregate data. For more detailed information see Table 1.

Table 1

Number of positive and negative data sets with category and type

Characteristics	Nr. of Studies	Positive	Negative
Clinical/Health			
Aggregate Data	-	-	-
Experiment	1	0	1 (100%)
Questionnaire	8	1 (13%)	7 (87%)
Consumer behavior			
Aggregate Data	-	-	-
Experiment	4	1 (25%)	3 (75%)
Questionnaire	3	1 (33%)	2 (66%)
Criminology			
Aggregate Data	1	0	1 (100%)
Experiment	1	0	1 (100%)
Questionnaire	-	-	-

(continued)

Characteristics	Nr. of Studies	Positive	Negative
Hospitality			
Aggregate Data	-	-	-
Experiment	2	1 (50%)	1 (50%)
Questionnaire	1	1 (100%)	0
Personality/Attitudes			
Aggregate Data	2	1 (50%)	1 (50%)
Experiment	-	-	-
Questionnaire	4	1 (25%)	3 (75%)
Pro. Decision Making			
Aggregate Data	-	-	-
Experiment	13	6 (46%)	7 (54%)
Questionnaire	4	3 (75%)	1 (25%)
Ratings			
Aggregate Data	-	-	-
Experiment	-	-	-
Questionnaire	3	1 (33%)	2 (66%)
Social Behavior			
Aggregate Data	-	-	-
Experiment	2	1 (50%)	1 (50%)
Questionnaire	-	-	-
Taxes			
Aggregate Data	1	1 (100%)	0
Experiment	1	1 (100%)	0
Questionnaire	-	-	-
Total			
Aggregate Data	4	2 (50%)	2 (50%)
Experiment	24	10 (42%)	14 (58%)
Questionnaire	23	8 (35%)	15 (65%)

Note. See Appendix B for research papers with categories.

There were no systematic differences between different methods or categories. Another closer look on all samples was made to find if studies which are positive or negative somehow differ systematically from other studies. A further comparison was made with all studies which provided more accurate data than just students and with a similarity to their non-student sample. As most studies didn't differentiate between what kind of students were used, this comparison ended up with a small sample size ($N=8$). Of those eight data sets there were no differences in five data sets between the student and the other sample, and in three there were significant differences between the two samples (see Table 2).

Table 2.

Student samples with similar non-student samples

Student sample	Non- student sample	No differences	Significant differences
Students with negotiation training	Negotiators	✓	
Students without training	Negotiators		✓
Advanced accounting students	Accountants	✓	
Graduate Business students	Managers	✓	
Graduate Business students	Managers	✓	
Undergraduate Business students	Managers		✓
Undergraduate accounting students	Accountants	✓	
Undergraduate business students	Managers		✓

Note. See Appendix B for research papers used in Table 2.

Discussion

Overall the results of this systematic review show that student samples are usually not appropriate surrogates for non-students as 61% of all data sets are negative, in contrast 39% are positive. Especially in the category *Clinical/health* some studies were not always designed to show if student data produced the same data when compared to a patient sample, but for factor structure of questionnaires or for other reasons. In other cases, students were compared to samples which were very different, such as prison samples. Due to the low sample sizes, the categories and the types didn't show a clear picture in what cases student sample data is comparable to non-student data. The low sample sizes also led to few

categories with only two studies. For example, in the case of social behavior one data set is positive and one is negative. As for taxes both data sets indicate that a student's behavior on tax compliance in a laboratory setting is comparable to a non-student sample in the same setting. In both categories the number of data sets is at the minimum for becoming a new category which was two. In research about tax compliance which was not obtained through this method by Choo, Fonesca and Myles (2014), students which did not pay taxes were compared with company employees whose income is reported by a third party, and self-employed taxpayers which were responsible for paying taxes themselves. Their results indicated that self-assessed taxpayers are the most compliant whereas students were the least compliant. According to the researchers this happened because of norms of compliance from outside of the laboratory. Those results about tax compliance are contradictory to the results obtained in this systematic review. For a clearer picture of categories and types, more studies are needed. As mentioned before students may be appropriate surrogates for non-students if they are similar in age, experience, education or knowledge to the sample which is being investigated (Briggs et al., 1996; Shuptrine, 1975; Jones & Sonner, 2001). Only very few studies drew direct a conclusion between two different student samples to a non-student sample, those studies will be discussed in detail. In a study by Remus (1996) graduate students and undergraduate students were compared to a sample of managers. In the experiment, participants were asked to make typical management decisions about how many workers to employ, the current workforce, size and inventory. The results showed that the standard errors of graduate students and managers were significantly lower than those of undergraduates. The experiment also showed that undergraduate students were more erratic in their decision making, made more costly decisions and used less effective heuristics. Just as graduate students received more training during their academic career than undergraduate students, another study indicated that there can be significant changes when students receive training. In a study by Herbst and Schwarz (2011) professional negotiators were compared to students without training and students which received negotiation training. The professionals were not significantly outperformed by the students which received negotiation training. Advanced accounting students (Mortensen, Fisher & Wines 2012) and undergraduate accounting students (Liyinarachi & Milne, 2005) made similar decisions as professional accountants. Graduate business students also made similarly good decisions as managers (Remus, 1986). Whereas undergraduate business students (South, 1974) and students which did not receive training (Herbst, 2011) did not make similarly good decisions as their professional sample. This shows that some student samples can compensate for other

populations where others won't. Those results indicate that the question isn't if student samples are appropriate, but which student samples are appropriate for which non-student population, and which variables they need to be similar on, such as age or experience.

More research needs to be done in what cases which students are appropriate for which other populations. It seems logical that a graduate business student is a better proband for a real manager than an undergraduate business student. In consumer behavior certain variables, such as age, distinguishing students from consumers need to be determined before conducting research with a student sample. It is important to point out that the use of student samples should not be rejected overall. Previous studies also have shown that students can serve as appropriate surrogates, depending on the population being studied (Fuchs & Sarstedt, 2009). Researchers need to be more cautious about which student sample they want to choose, when the student sample is just a replacement for the real population of interest. For future research it would be desirable for researchers to be more specific about which kind of students were used.

Limitations

The first limitation of this work is the kind of literature review which was used. The systematic quantitative approach can only include studies which were found through the method as previously described, whereas in a typical review all kinds of studies can be included. Problematic is that the typical approach is also prone to several biases, as the researcher can choose themselves which studies to include or exclude (Petticrew, 2001). As the problem if student samples are comparable to other populations isn't being researched as much as it should be, this systematic review ended up with a small sample size. The weakness of this method's approach is being compensated by an introduction and discussion which includes research which was obtained nonetheless of the method. The biggest limitation of this work is the small sample size. This small sample size didn't allow to draw conclusions between categories or types of studies. As previously mentioned many researchers didn't seem to recognize the problems which arise with student samples, so they didn't specify what kind of students were used. So, in most cases there couldn't be made any more precise comparisons including type and category. There seem to be differences between how advanced a student is like graduate or undergraduate students, which course a student is in, in a student's experience with a task and other variables such as age. In only a few studies

the researchers differentiated between graduate students or undergraduate students, or other kinds of students. Some studies had to be excluded because researchers mixed up student samples with non-students, usually university staff, and didn't differentiate between them in their statistical analysis. Those limitations show that more research on when students samples, or under what circumstances, are appropriate surrogates for non-student populations needs to be done, so these gaps can be filled.

References

- Ashton, R. H., & Kramer, S. S. (1980). Students as surrogates in behavioral accounting research: Some evidence. *Journal of Accounting Research*, 1-15.
- Ashraf, R., & Merunka, D. (2017). The use and misuse of student samples: An empirical investigation of European marketing research. *Journal of Consumer Behaviour*, 16(4), 295-308.
- Barr, S. H., & Hitt, M. A. (1986). A comparison of selection decision models in manager versus student samples. *Personnel Psychology*, 39(3), 599-617.
- Briggs, R. O., Balthazard, P. A., & Dennis, A. R. (1996). Graduate business students as surrogates for executives in the evaluation of technology. *Journal of Organizational and End User Computing*, 8(4), 11-19.
- Choo, L., Fonseca, M. A., & Myles, G. D. (2014). Do students behave like real taxpayers? Experimental evidence on taxpayer compliance from the lab and from the field. *Tax Administration Research Centre* 4, 14.
- Cook, T. D., Campbell, D. T., & Shadish, W. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston: Houghton Mifflin.
- Corrigan, J. R., Depositario, D. P. T., Nayga Jr, R. M., Wu, X., & Laude, T. P. (2009). Comparing open-ended choice experiments and experimental auctions: an application to golden rice. *American Journal of Agricultural Economics*, 91(3), 837-853.
- Depositario, D. P. T., Nayga Jr, R. M., Wu, X., & Laude, T. P. (2009). Should students be used as subjects in experimental auctions? *Economics Letters*, 102(2), 122-124.
- Espinosa, J. A., & Ortinau, D. J. (2016). Debunking legendary beliefs about student samples in marketing research. *Journal of Business Research*, 69(8), 3149-3158.
- Fuchs, S., & Sarstedt, M. (2009). Is there a tacit acceptance of student samples in marketing and management research? *International Journal of Data Analysis Techniques and Strategies*, 2(1), 62-72.
- Goode William, J. (1952). *Methods in Social Research*. McGraw-Hill Book Company; New York; Toronto; London.

- Henggeler, S. W., Sallis, J. F., & Cooper, P. F. (1980). Comparison of university mental health needs priorities identified by professionals and students. *Journal of Counseling Psychology*, 27(2), 217.
- Herbst, U., & Schwarz, S. (2011). How valid is negotiation research based on student sample groups? New insights into a long-standing controversy. *Negotiation Journal*, 27(2), 147-170.
- Hooghe, M., Stolle, D., Mahéo, V. A., & Vissers, S. (2010). Why can't a student be more like an average person? Sampling and attrition effects in social science field and laboratory experiments. *The Annals of the American Academy of Political and Social Science*, 628(1), 85-96.
- Jones, W. L., & Sonner, B. S. (2001). Just say no to traditional student samples. *Journal of Advertising Research*, 41(5), 63-71.
- Kardes, F. R. (1996). In defense of experimental consumer psychology. *Journal of Consumer Psychology*, 5(3), 279-296.
- Landrum, R. E., & Chastain, G. (1995). Experiment spot-checks: A method for assessing the educational value of undergraduate participation in research. *IRB: Ethics & Human Research*, 17(4), 4-6.
- Liyanarachchi, G. A., & Milne, M. J. (2005). Comparing the investment decisions of accounting practitioners and students: an empirical study on the adequacy of student surrogates. *Accounting Forum Vol. 29*, 121-135
- Lucas, J. W. (2003). Theory-testing, generalization, and the problem of external validity. *Sociological Theory*, 21(3), 236-253.
- Lynch Jr, J. G. (1982). On the external validity of experiments in consumer research. *Journal of consumer Research*, 9(3), 225-239.
- Megehee, C. M. (2009). Advertising time expansion, compression, and cognitive processing influences on consumer acceptance of message and brand. *Journal of Business Research*, 62(4), 420-431.
- Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group (2009) Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. *PLoS Medicine* 6(7), e1000097.

- Mook, D. G. (1983). In defense of external invalidity. *American psychologist*, 38(4), 379.
- Mortensen, T., Fisher, R., & Wines, G. (2012, December). Students as surrogates for practicing accountants: Further evidence. *Accounting Forum* 36, 4, 251-265.
- Muehlbacher, S., & Kirchler, E. (2016). Taxperiments. About the external validity of laboratory experiments in tax compliance research. *Die Betriebswirtschaft (DBW)*, 76, 7-19.
- Newman, D. A., Joseph, D. L., & Feitosa, J. (2015). External validity and multi-organization samples: Levels-of-analysis implications of crowdsourcing and college student samples. *Industrial and Organizational Psychology*, 8(2), 214-220.
- Park, C. W., & Lessig, V. P. (1977). Students and housewives: Differences in susceptibility to reference group influence. *Journal of consumer Research*, 4(2), 102-110.
- Peterson, R. A. (2001). On the use of college students in social science research: Insights from a second-order meta-analysis. *Journal of Consumer Research*, 28, 450-461.
- Peterson, R. A., & Merunka, D. R. (2014). Convenience samples of college students and research reproducibility. *Journal of Business Research*, 67(5), 1035-1041.
- Petticrew, M. (2001). Systematic reviews from astronomy to zoology: myths and misconceptions. *The BMJ*, 322(7278), 98-101.
- Pickering, C., & Byrne, J. (2013). The benefits of publishing systematic quantitative literature reviews for PhD candidates and other early-career researchers. *Higher Education Research & Development*, 33(3), 534-548.
- Remus, W. (1996). Will behavioral research on managerial decision making generalize to managers? *Managerial and Decision Economics*, 93-101.
- 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(3), 515.
- Shuptrine, F. K. (1975). On the validity of using students as subjects in consumer behavior investigations. *The Journal of Business*, 48(3), 383-390.
- Simonson, I., Carmon, Z., Dhar, R., Drolet, A., & Nowlis, S. M. (2001). Consumer research: In search of identity. In S. T. Fiske, D. L. Schacter, & C. Zahn-Waxler (Eds.), *Annual Review of Psychology*, Vol. 52. 249-275

South, J. C. (1974). Achievement motivation among managers of small businesses, corporation managers, and business students. *Journal of Applied Psychology*, 59(4), 509.

Thomas, R. W. (2011). When student samples make sense in logistics research. *Journal of Business Logistics*, 32(3), 287-290.

List of Figures

Figure 1. Prisma Flow Diagram.....	9
Figure 2. Fifteen stages in undertaking systematic quantitative literature reviews.....	10

List of Tables

Table 1. Number of positive and negative data sets with category and type.....	15
Table 2. Student samples with similar non-student sample.....	17

Appendix

Appendix A

Abstract

Student samples are the sample of choice in many psychological studies. This is not because students are the actual population of interest, but because they are readily available or the result of convenience sampling. This approach might give rise to problems with reliability and external validity. This systematic review analyzes literature which compares student samples to non-student samples and draws conclusions if those samples produce the same results. Overall of 51 data sets 31 show that student samples are not appropriate surrogates for non-student samples, and in 20 cases student samples are appropriate. More specifically it shows it depends much more on which kind of student is used for which non-student sample such as a graduate business student is suitable for a manager. Also, it shows that overall most researchers don't recognize a problem with using student samples and don't differentiate between different students or how advanced they are.

Keywords: student samples, student surrogates, external validity

Zusammenfassung

Psychologische Forschung benutzt öfters Studenten als Stichprobe. Dies passiert nicht, weil Studenten jene Population sind welche erforscht wird, sondern weil Studenten leicht zu Forschungszwecken verfügbar sind. Diese Herangehensweise kann zu Problemen mit Reliabilität und externer Validität führen. Dieses systematische Review analysiert Literatur, welche Daten von studentischen Stichproben mit Daten von Nicht-Studenten vergleicht, und zieht Konklusionen, ob diese Stichproben zu denselben Ergebnissen führen. Insgesamt zeigen 31 von 51 Datensätzen, dass studentische Stichproben kein Ersatz für Nicht-Studenten sind, wohingegen 20 Datensätze zeigen, dass Daten von beiden Stichproben zu denselben Ergebnissen führen. Genauer betrachtet zeigt sich, dass es von Relevanz ist, welche Studenten für welche Nicht-Studenten Stichproben benutzt werden, wie beispielsweise ein fortgeschrittener Betriebswirtschaftsstudent für einen Manager. Weiters zeigt sich, dass in der Forschung das Problem mit studentischen Stichproben oft nicht erkannt wird, und Forscher oft nicht zwischen verschiedenen Studenten unterscheiden oder wie weit fortgeschritten diese in ihrem Studium sind.

Keywords: student samples, student surrogates, external validity

Appendix B

Professional Decision Making

- Ashton, R. H., & Kramer, S. S. (1980). Students as surrogates in behavioral accounting research: Some evidence. *Journal of Accounting Research*, 1-15.
- Barr, S. H., & Hitt, M. A. (1986). A comparison of selection decision models in manager versus student samples. *Personnel Psychology*, 39(3), 599-617.
- Chen, C. C., Schwartz, Z., & Yu, X. J. (2015). On the 'students as surrogates' research conundrum: Insights from statistical reasoning in revenue management. *Journal of Revenue and Pricing Management*, 14(4), 262-275.
- Herbst, U., & Schwarz, S. (2011). How valid is negotiation research based on student sample groups? New insights into a long-standing controversy. *Negotiation Journal*, 27(2), 147-170.
- Hughes, C. T., & Gibson, M. L. (1991). Students as surrogates for managers in a decision-making environment: An experimental study. *Journal of Management Information Systems*, 8(2), 153-166.
- Lehmann, E. L. (2012). "Student" and small-sample theory. In Selected works of EL Lehmann (pp. 1001-1008). Springer, Boston, MA.
- Liyanarachchi, G. A., & Milne, M. J. (2005). Comparing the investment decisions of accounting practitioners and students: an empirical study on the adequacy of student surrogates. *Elsevier. In Accounting Forum Vol. 29*, 121-135 Elsevier.
- Mintz, A., Redd, S. B., & Vedlitz, A. (2006). Can we generalize from student experiments to the real world in political science, military affairs, and international relations? *Journal of Conflict Resolution*, 50(5), 757-776.
- Mortensen, T., Fisher, R., & Wines, G. (2012, December). Students as surrogates for practicing accountants: Further evidence. *Accounting Forum* 36, 4, 251-265.
- Porter, T. L., & Cook, T. E. (1964). A comparison of student and professional prestige ranking of jobs in psychology. *Journal of Counseling Psychology*, 11(4), 385.

Remus, W. (1986). Graduate students as surrogates for managers in experiments on business decision making. *Journal of Business Research*, 14(1), 19-25.

Remus, W. (1996). Will behavioral research on managerial decision making generalize to managers? *Managerial and Decision Economics*, 93-101.

Singer, M. (1990). Implicit leadership theory: are results generalizable from student to professional samples? *The Journal of social psychology*, 130(3), 407-408.

Trottier, K., & Gordon, I. M. (2018). Students as surrogates for managers: Evidence from a replicated experiment. *Canadian Journal of Administrative Sciences/Revue Canadienne des Sciences de l'Administration*, 35(1), 146-161.

Waldron, M., & Fisher, R. (2017). Values and ethical judgments: The adequacy of students as surrogates for professional accountants. *Meditari Accountancy Research*, 25(1), 37-64.

Social behavior

Anderson, J., Burks, S. V., Carpenter, J., Götte, L., Maurer, K., Nosenzo, D. & Rustichini, A. (2013). Self-selection and variations in the laboratory measurement of other-regarding preferences across subject pools: evidence from one college student and two adult samples. *Experimental Economics*, 16(2), 170-189.

Falk, A., Meier, S., & Zehnder, C. (2013). Do lab experiments misrepresent social preferences? The case of self-selected student samples. *Journal of the European Economic Association*, 11(4), 839-852.

Clinical health

Balogh, D. W., & Merritt, R. D. (1990). Accounting for schizophrenics' Magical Ideation scores: Are college-student norms relevant?

Barker-Collo, S., & Moskowitz, A. K. (2005). Profiles of DES performance in inmate and student samples. *Journal of trauma & dissociation*, 6(1), 113-124.

Chan, C. H. Y., Chan, T. H. Y., Peterson, B. D., Lampic, C., & Tam, M. Y. J. (2014). Intentions and attitudes towards parenthood and fertility awareness among Chinese university students in Hong Kong: a comparison with Western samples. *Human Reproduction*, 30(2), 364-372.

- Cox, B. J., Enns, M. W., & Larsen, D. K. (2001). The continuity of depression symptoms: Use of cluster analysis for profile identification in patient and student samples. *Journal of affective disorders*, 65(1), 67-73.
- Crössmann, A., & Pauli, P. (2006). The factor structure and reliability of the Illness Attitude Scales in a student and a patient sample. *BMC psychiatry*, 6(1), 46.
- Hemming, J. H. (1981). Electrodermal indices in a selected prison sample and students. *Personality and Individual Differences*, 2(1), 37-46.
- Patel, T., Mansell, W., & Veale, D. (2015). The cognitive behavioural processes questionnaire: a preliminary analysis within student, mixed clinical and community samples and the identification of a core transdiagnostic process. *Cognitive Therapy and Research*, 39(2), 193-203.
- Pompili, M., Erbuto, D., Innamorati, M., Migliorati, M., & Girardi, P. (2017). Assessment of suicide-related knowledge and skills in a sample of health professionals and students. *European Psychiatry*, 41, S300-S301.
- Stiefel, F., Krenz, S., Zdrojewski, C., Stagno, D., Fernandez, M., Bauer, J. & Fegg, M. (2008). Meaning in life assessed with the “Schedule for Meaning in Life Evaluation” (SMiLE): a comparison between a cancer patient and student sample. *Supportive Care in Cancer*, 16(10), 1151-1155.

Consumer behavior

- Beltramini, R. F. (1983). Student surrogates in consumer research. *Journal of the Academy of Marketing Science*, 11(4), 438-443.
- Burnett, J. J., & Dune, P. M. (1986). An appraisal of the use of student subjects in marketing research. *Journal of Business Research*, 14(4), 329-343.
- Jones, W. L., & Sonner, B. S. (2001). Just say no to traditional student samples. *Journal of Advertising Research*, 41(5), 63-71.
- Morgan, F. W. (1979). Students in marketing research: Surrogates vs. role-players. *Journal of the Academy of Marketing Science*, 7(3), 255-264.
- Park, C. W., & Lessig, V. P. (1977). Students and housewives: Differences in susceptibility to reference group influence. *Journal of consumer Research*, 4(2), 102-110.

Tih, S., Ennis, S., & Poon, J. M. (2008). Attitude towards online retailing services: A comparison of student and non-student samples. *Asian Academy of Management Journal of Accounting and Finance*, 13(1), 113-126.

Ratings

Cleveland, J. N., & Berman, A. H. (1987). Age perceptions of jobs: Agreement between samples of students and managers. *Psychological Reports*, 61(2), 565-566.

Doyle, K. O., & Crichton, L. I. (1978). Student, peer, and self-evaluations of college instructors. *Journal of Educational Psychology*, 70(5), 815.

Henggeler, S. W., Sallis, J. F., & Cooper, P. F. (1980). Comparison of university mental health needs priorities identified by professionals and students. *Journal of Counseling Psychology*, 27(2), 217.

Criminology

Brady, P. Q., Nobles, M. R., & Bouffard, L. A. (2017). Are college students really at a higher risk for stalking? Exploring the generalizability of student samples in victimization research. *Journal of criminal justice*, 52, 12-21.

Smart, S. M., Berry, M. A., & Rodriguez, D. N. (2014). Skilled observation and change blindness: A comparison of law enforcement and student samples. *Applied Cognitive Psychology*, 28(4), 590-596.x

Personality/Attitudes

Flere, S., & Lavrič, M. (2008). On the validity of cross-cultural social studies using student samples. *Field methods*, 20(4), 399-412.

Hanel, P. H., & Vione, K. C. (2016). Do Student Samples Provide an Accurate Estimate of the General Public? *PloS one*, 11(12), e0168354.

Shen, H. (2017). Refining organization–public relationship quality measurement in student and employee samples. *Journalism & Mass Communication Quarterly*, 94(4), 994-1010.

South, J. C. (1974). Achievement motivation among managers of small businesses, corporation managers, and business students. *Journal of Applied Psychology*, 59(4), 509.

Wood, E. F., & Miller, M. K. (2016). Prejudice toward immigrants in student and community samples. *Journal of Aggression, Conflict and Peace Research*, 8(4), 290-300.

Hospitality

Ok, C., Shanklin, C. W., & Back, K. J. (2008). Generalizing survey results from student samples: implications from service recovery research. *Journal of quality assurance in hospitality & tourism*, 8(4), 1-23.

Raybould, M., & Wilkins, H. (2006). Generic skills for hospitality management: A comparative study of management expectations and student perceptions. *Journal of Hospitality and Tourism Management*, 13(2), 177-188.

Ro, H., & Kubickova, M. (2013). The use of student subjects in hospitality research: insights from subjective knowledge. *Journal of Quality Assurance in Hospitality & Tourism*, 14(4), 295-320.

Taxes

Alm, J. & McKee, M. (2011). Comparing student and non-student reporting behavior in tax compliance experiments. *Recent Research on Tax Administration and Compliance*, 93.

Alm, J., Cherry, T. L., Jones, M., & McKee, M. (2012). Social programs as positive inducements for tax participation. *Journal of Economic Behavior & Organization*, 84(1), 85-96.

Student samples with similar non-student samples

Herbst, U., & Schwarz, S. (2011). How valid is negotiation research based on student sample groups? New insights into a long-standing controversy. *Negotiation Journal*, 27(2), 147-170.

Liyanarachchi, G. A., & Milne, M. J. (2005). Comparing the investment decisions of accounting practitioners and students: an empirical study on the adequacy of student surrogates. *Elsevier. In Accounting Forum Vol. 29*, 121-135 Elsevier.

Mortensen, T., Fisher, R., & Wines, G. (2012, December). Students as surrogates for practicing accountants: Further evidence. *Accounting Forum* 36, 4, 251-265.

- Remus, W. (1986). Graduate students as surrogates for managers in experiments on business decision making. *Journal of Business Research*, 14(1), 19-25.
- Remus, W. (1996). Will behavioral research on managerial decision making generalize to managers? *Managerial and Decision Economics*, 93-101.
- South, J. C. (1974). Achievement motivation among managers of small businesses, corporation managers, and business students. *Journal of Applied Psychology*, 59(4), 509.