STATISTICAL APPROACHES FOR DOCTORAL PROJECTS
PRONESLLC. ©

We will help this much!

Contents

DIFFERENCES BETWEEN QUALITY IMPROVEMENT AND HUMAN SUBJECTS RESEARCH

Before you plan and conduct an investigational project, you must determine if the planned activities to be employed are human subjects research or a quality improvement project (Bass & Maloy, 2020). Specific regulations like the Code of Federal Regulations and Health and Human Services govern the conduct of human subjects research. When the investigator is making this determination, it can be confusing because human subjects research and quality improvement projects share similar characteristics.

While these investigations are both rigorous processes and at times involve similar methods, the two types of investigations have distinctly different overall aims. Quality improvement projects use *data-driven methods* to improve health delivery and quality. Such projects examine changes in human behavior and are largely experiential learning processes associated with the sample and the setting involved. The patient is not an active subject in a quality improvement project but instead the data resulting from their care is examined for any change or improvement. The results cannot be generalized beyond the setting and sample selected in the quality improvement project.

In contrast, research is a systematic investigation designed to develop or contribute to generalizable knowledge which may involve procedures requiring the consent of the patient or participant. The procedures associated with the research are not considered patient care but research measures. Further, visits to the clinic associated with a research project may require consideration as to whether it is even patient care but instead being conducted primarily for research procedures. In this case, should a patient even be charged for the visit if it is research.

This article below provides good insight into the differences associated with the two approaches. It is important that the manuscript and subsequent publications make these approaches clear and distinct.

Bass, P. F., & Maloy, J. W. (2020). How to Determine if a Project Is Human Subjects Research, a Quality Improvement Project, or Both. *The Ochsner journal*, *20*(1), 56–61. https://doi.org/10.31486/toj.19.0087

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7122260/pdf/TOJ-19-0087_56Bass.pdf

COMPARISON OF RESEARCH, EVIDENCE BASED PRACTICE & QUALITY IMPROVEMENT

	RESEARCH	EVIDENCE BASED PRACTICE	QUALITY IMPROVEMENT
	Research Generates Evidence. Evidence used for practice.	Evidence Based Practice involves translating and implementing the Research Evidence within a setting.	Quality Improvement involves sustaining Best Practice within a setting.
Question:	*What can we do?*	*Are we using a best practice?*	*Are we using a best practice, doing it right, and continuously?*
Focus:	▪ Research discovers new knowledge or theory that be generalized to populations.	▪ Evidence based practice translates evidence into practice by implementing it in specific settings and measuring the outcomes.	▪ Quality improvement generates internal evidence about a practice and the process outcomes for specific populations/settings.
Purpose:	▪ Research intends to generate knowledge that can be used in various settings, including the laboratory, clinical trials, and epidemiology.	▪ The evidence is implemented into an institutional practice to impact processes and outcomes. The evidence has not been used previously.	▪ Quality improvement sustains and improves a best practice to directly impact institutional outcomes.
Evidence	▪ The process involves examining the literature to determine the gap, determine a research question, test a hypothesis, perform a valid and reliable assessment, and generate, disseminate, and generalize the findings.	▪ Appraise the evidence internally at the site and in the literature to determine the best practice. Clinical expertise, patient preference and values may influence the choice. Implement the recommendations and modify them based on feedback. Measure outcomes and disseminate the findings locally and regionally.	▪ Quality improvement monitors evidence based interventions and initiatives as well as outcomes continuously. The process is to implement strategies which will be sustained to continue generating internal evidence. Disseminates evidence as approved by the setting.
Methods	▪ The methods are quantitative, qualitative, and mixed.	▪ Methods can involve research methods, 7 steps of evidence based practice, the IOWA model, Advancing Research and Clinical practice through close Collaboration (ARCC) Model	▪ Methods can involve research methods or involve Plan, Do, Study, Act, Define, measure, analyze, improve, and control (DMAIC), LEAN, Six Sigma, Total Quality Management
Approval	▪ Requires the consideration of an institutional review board if it involves humans or living animals. Human subject protection is required.	▪ For the protection of data and publication requirements, an Institutional review board or other internal review is required.	▪ For protection of data and publication requirements, Institutional review board or other internal review required.

These are the four types of quantitative designs. Depending on the scope of the project or study, one of these will be selected for conduct. Your program may want you to use Quality Improvement Project to avoid sounding like research.

Figure 1

QUANTITATIVE, QUASI-EXPERIMENTAL DESIGN – AKA COMPARATIVE DESIGN

WHAT TYPE OF DESIGN IS MY DIRECT PRACTICE IMPROVEMENT PROJECT?
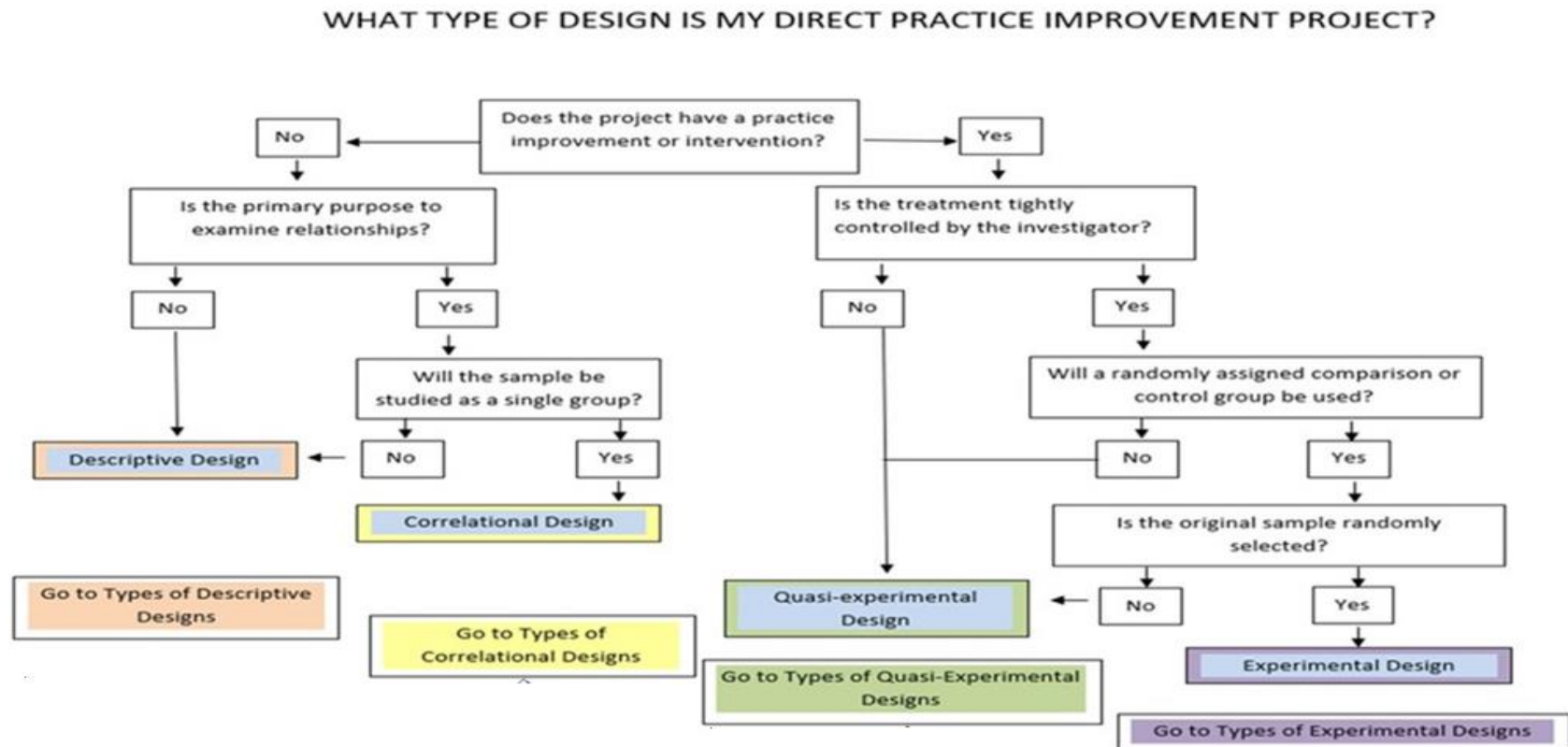
The type of quasi-experimental or causal comparative design depends on the type of sampling. Two types of sampling are within group (paired) or between group (independent). These are demonstrated in the next diagram.

Figure 2

Quasi-experimental Design – Causal Comparative Design

QUASI-EXPERIMENTAL DESIGN

Types of Quasi-Experimental Designs

Does the Quasi Experimental Project have a comparison group?

No

Yes

WITHIN GROUP COMPARISON

BETWEEN GROUP COMPARISON

Comparison Group versus Intervention Group

DPI
Write
Quantitative
Quasi-experimental

Also known as…

Causal Comparative
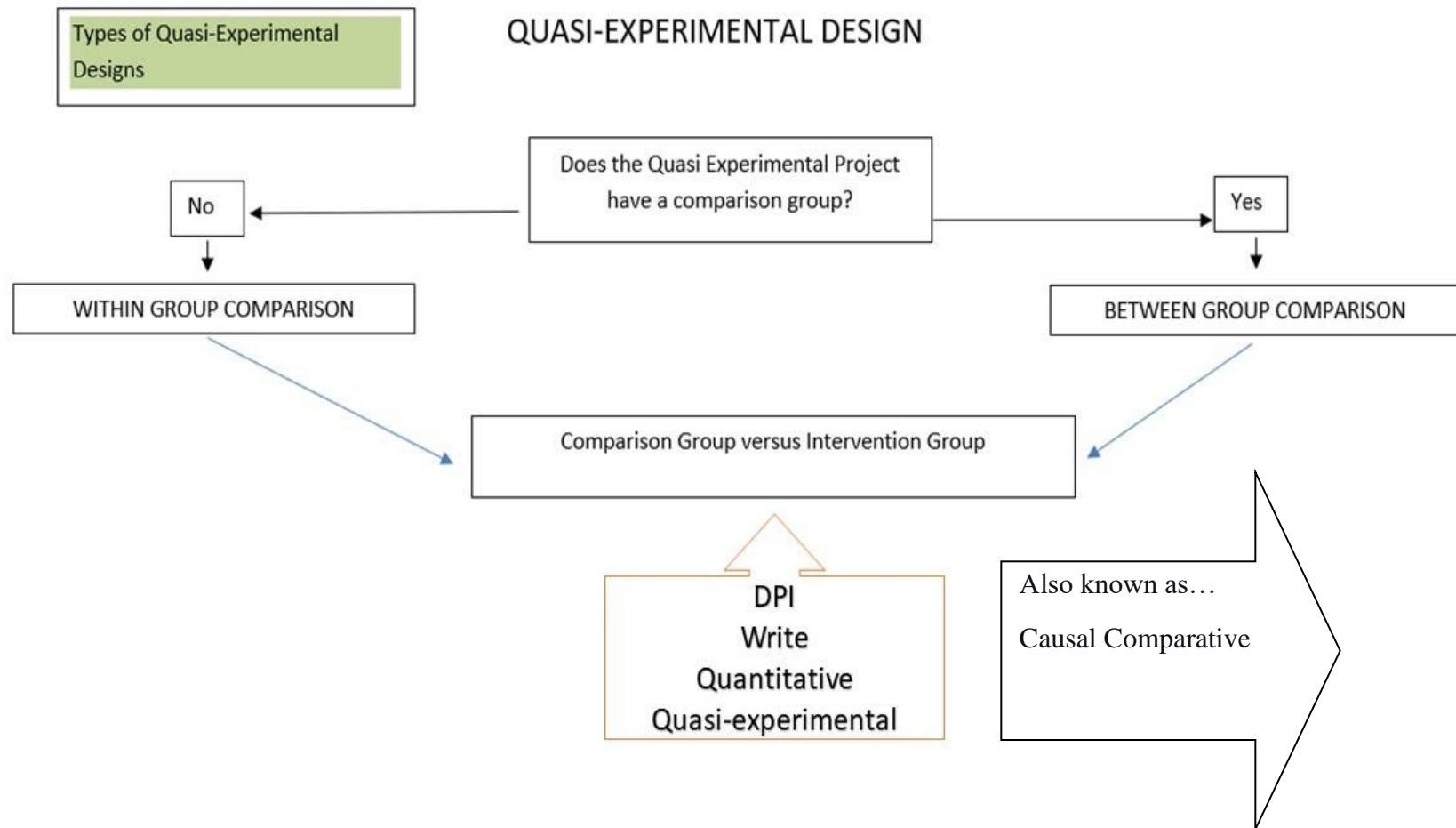
The within group is collecting data on the same group of people before and after an intervention. The between group is collecting data in two separate or non-equivalent groups

Figure 3

Type of Sampling: Between or Within Group
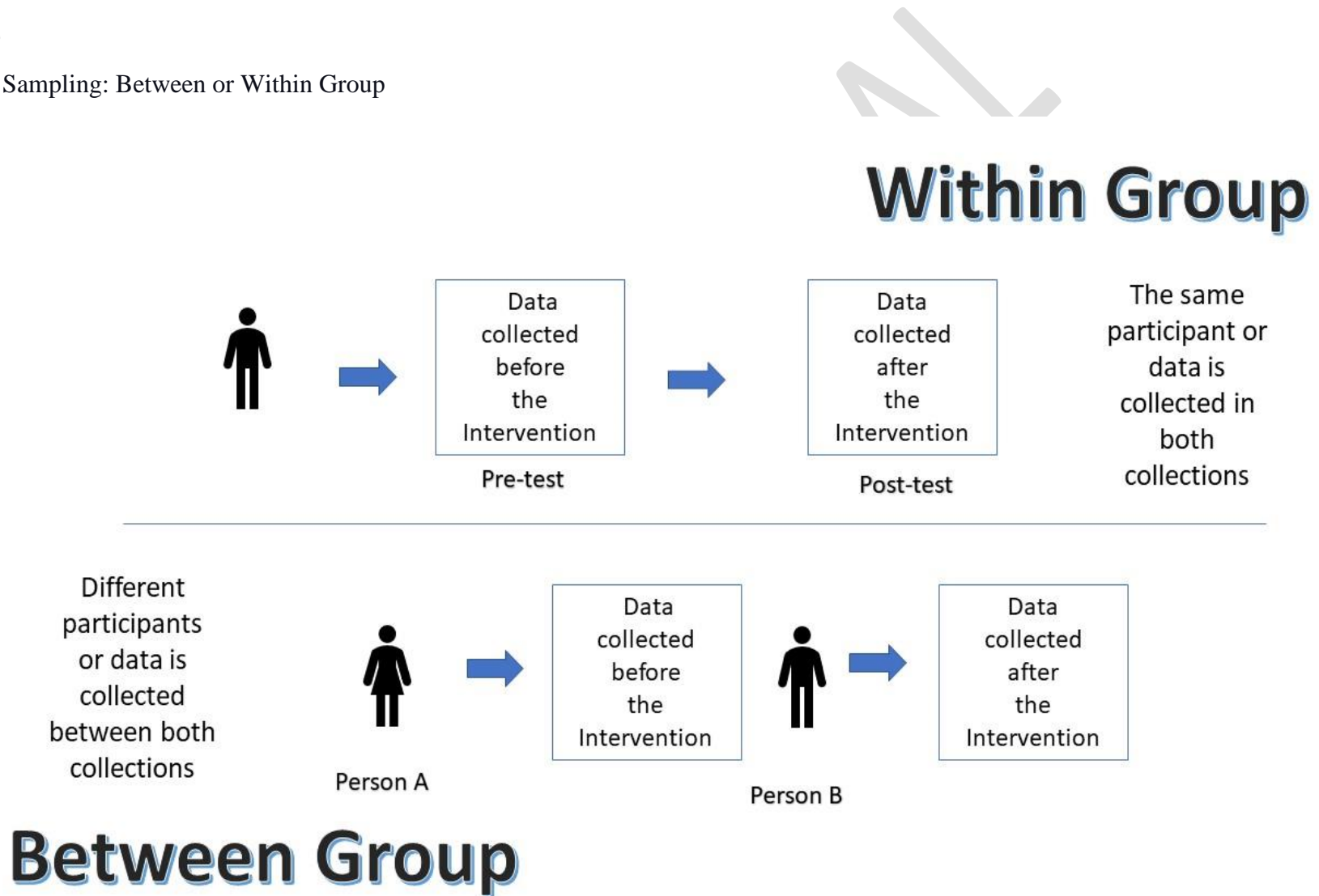
A PRIORI POWER ANALYSIS

Using the G*Power program, an estimation of the sample size can be conducted with predetermined levels of significance, power, and effect. A significance level of $p < .05$ was identified for statistical significance for the chi-square to be used for determining significance in the sepsis diagnosis rate. In the first power analysis, an a priori test for a chi-square was conducted using a significance of .05, power of .8 and large effect of 0.5. This test shows a recommended sample size of 32 or ideally 16 patients in each group.

Figure 4

Power Analysis For Chi-Square Analysis
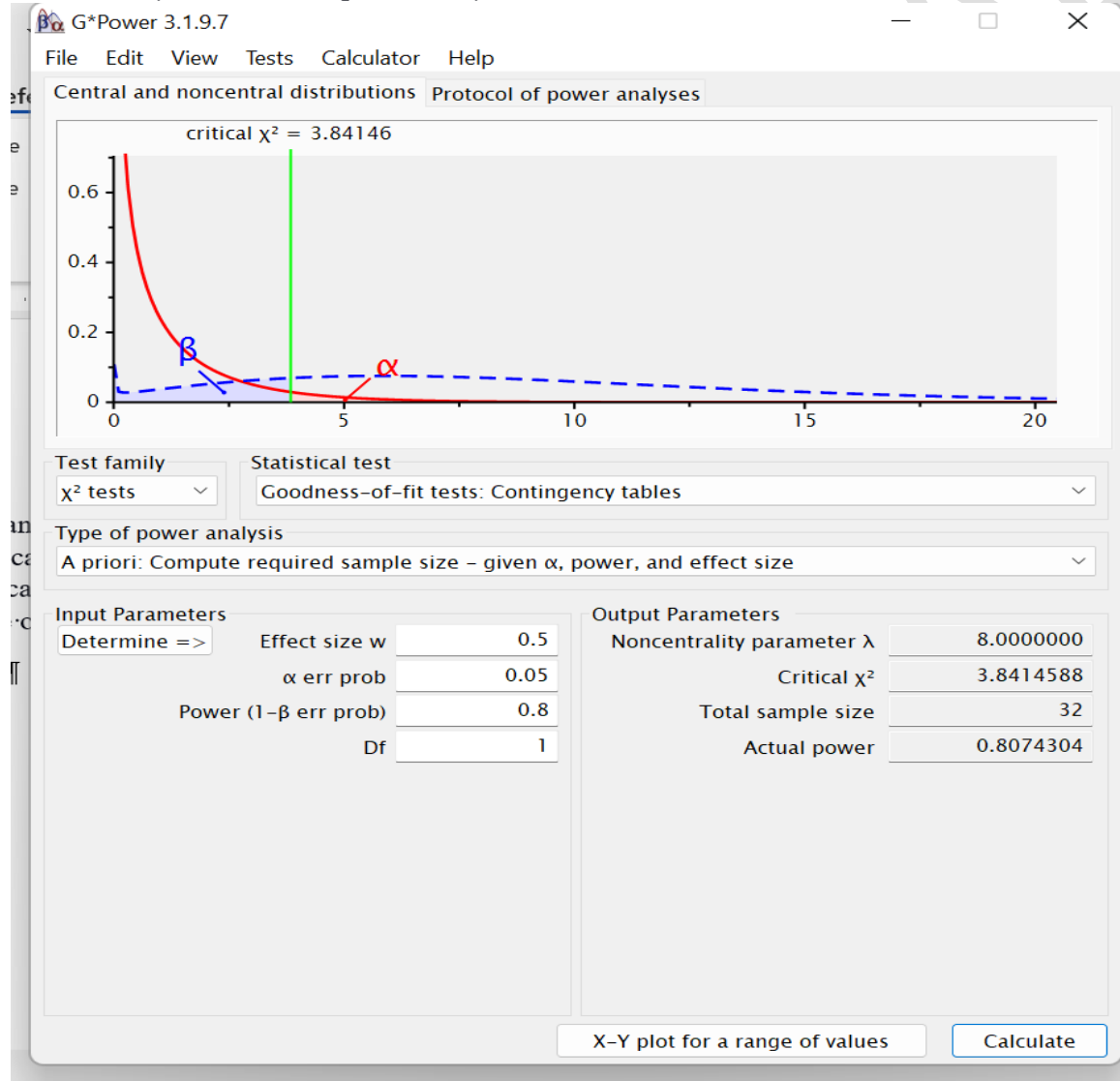
In the second power analysis, an a priori test for a chi-square was conducted using a significance of .05, power of .95 and large effect of 0.5. This test shows a recommended sample size of 52 or

ideally 26 patients in each group. The second analysis has a higher power, which makes the statistics more sensiive to change but also requires a bigger sample.

Figure 5

Power Analysis for Chi-Square Test
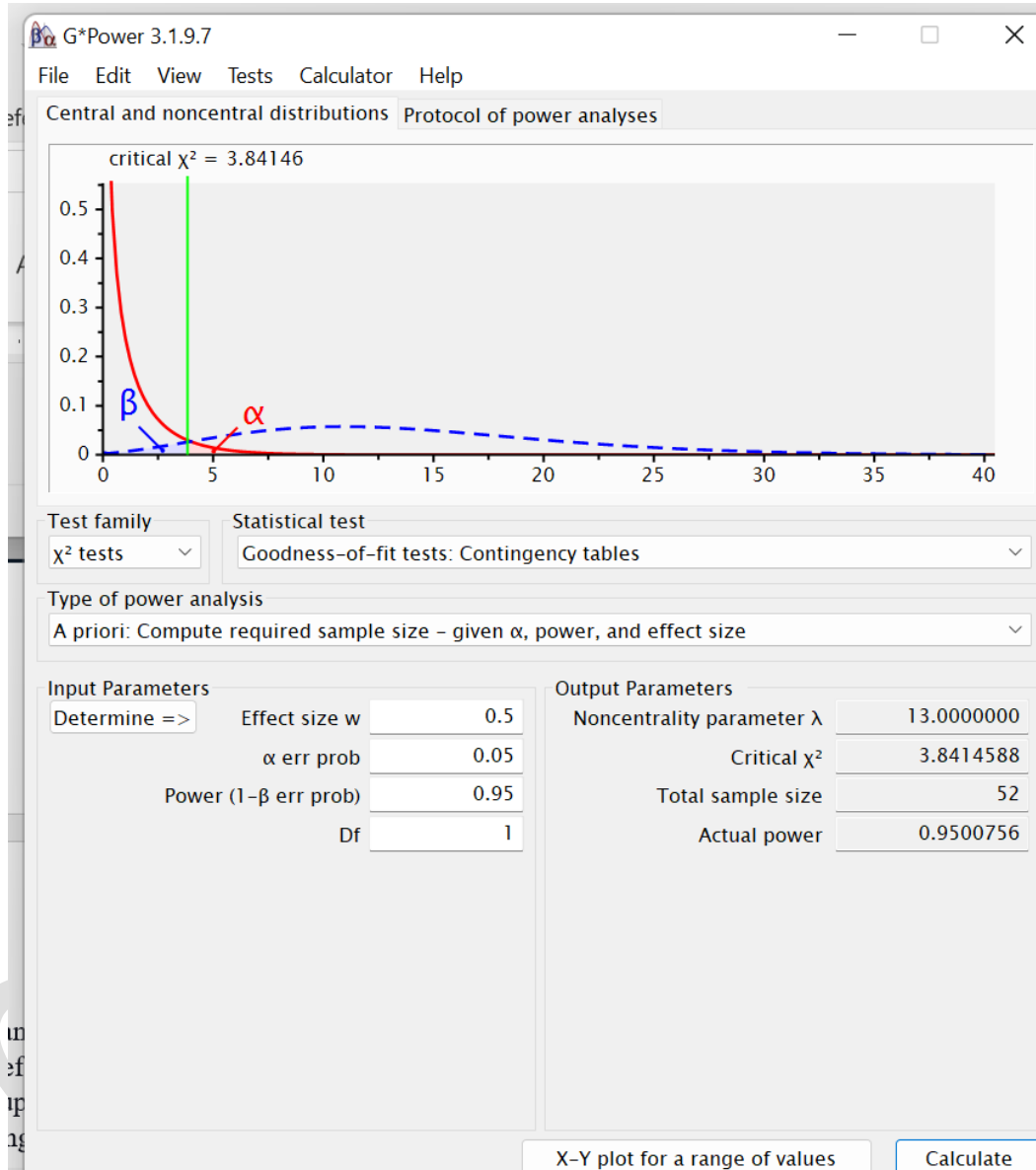
A post hoc power analysis can be calculated after the project has collected data to determine if the true sample size, significance level, and effect size were at the same level of power we used to establish the sample size. In the event that the outcome is not significant, a lack of power can be one factor creating a lack of significance.

PLAN FOR DATA ANALYSIS

1. Clearly states the clinical question for the DNP project or the research objectives and hypothesis for the Ph.D. thesis.
2. Identifies the dataset used, e.g., Microsoft Excel or other comma delimited text, to collect and prepare the data for analysis.
3. Share any Inclusion and exclusion criteria of data that will or will not be included and why.
4. State the independent and dependent variables as defined in the manuscript. Identify their level of measurement, including any coding used for nominal or ordinal levels. Coding is used to provide a numerical value for a variable which is a category or count. For example, males would be coded to 1 and females coded 2 because the statistical program cannot analyze the text male or female. Instead it analyzes the 1s and 2s and nominal level data.
5. State the desired outcomes of the clinical questions for DNP projects or Ph.D. research hypotheses using the planned statistical tests.
6. Indicates the software for statistical analysis e.g. SPSS version 28.
7. Identify any missing data, outliers and determine plan for use or removal. Plan to discuss the background for any issues.
8. Describe the sample using demographic variables e.g. age, gender, diagnosis. These are reported in a table using means, standard deviations or counts and percentages.
9. Identify the statistical tests based on the variables' levels of measurement.

   Parametric Tests using within group or between group t tests. If fails assumption testing, can report Wilcoxon Signed Rank or Mann Whitney U.

   Non-parametric Tests for between group involve the Pearson Chi-square test and Fishers Exact test. Within group comparisons use McNemars Chi-square test.

   Multivariate analyses may involve analysis of variance, Kruskal Wallis, multiple and logistic regression.
10. Creating shell tables and planned figures of results using APA style.
11. Write a narrative that addresses all entries in the tables and figures in the text. Both should stand alone.

COMMON STATISTICAL TESTS

Figure 5

Parametric Statistics for Analysis of Ratio or Interval Level Dependent Variable

Parametric Tests – Test where the Dependent Variable is an Interval or Ratio Level of Measurement					
Test Statistic	Purpose	Level of Measurement IV	Level of Measurement DV	Cov	Index of Strength for the Relationship
One sample t test (t)	to test for the predicted value of a mean to the population	-	Interval, Ratio	-	-
t-test for two independent groups (t)	to test the difference between the means of 2 independent groups	Nominal	Interval, Ratio	-	point bi-serial
t-test for two dependent groups (pre-test/post-test) (t)	to test the difference between the means of 2 related groups/sets of	Nominal	Interval, Ratio	-	point bi-serial
Analysis of Variance (ANOVA) (F)	to test the difference among the means of 3 or more independent groups (one-way) or froups with 2 + dependent variables (multi-way)	Nominal	Interval, Ratio	-	eta2
Repeated ANOVA (F)	to test the difference among 3+ related groups or sets of scores	Nominal	Interval, Ratio	-	eta2
Multivariate analysis of variance (MANOVA)	to test the difference between the means of 2+ groups for 2+ DV simultaneously	Nominal	Interval, Ratio	Nominal, Interval, Ratio	eta2
Analysis of Co-variance (ANCOVA)	to test the difference between the means of 2+ groups, while controlling for 1 co-variates	Nominal	Interval, Ratio	Nominal, Interval, Ratio	eta2
Multivariate analysis of covariance	to test the difference between the means of 2+ DV simultaneously while controlling about 1+ covariate	Nominal	Interval, Ratio	Nominal, Interval, Ratio	eta2
Pearson-product moment correlation	to test the relationship or correlation between two variables.	Interval, Ratio	Interval, Ratio		r2
Multiple correlation/regression	to test the relationship between two IV and 1DV; to predict a DV from 2+ IV	Nominal, Interval, Ratio	Interval, Ratio		r2
Logistic Regression	to test the relationship between 2+ IV and 1DV; to predict he probability of an event; to estimate relative risk	Nominal, Interval, Ratio	Nominal		r2

Note. Image taken from Creswell and Creswell (2018).

The independent variable within a quasi-experimental design will be a nominal or categorical level variable identifying the sample or group associated with the intervention. It is the dependent variable's level of measurement which will direct the type of statistical analysis e.g. parametric versus non-parametric. If the dependent variable is a ratio, interval, the test to be used would be a parametric one. If the dependent variable is an ordinal or nominal level, a non-parametric test would be used.

Figure 6

Non-Parametric Statistics for Analysis of Nominal or Ordinal Level Dependent Variable

Non- parametric Tests - Test where the Dependent Variable is an ordinal or nominal Level of Measurement				
		Level of Measurement		Index of Strength for the
Test Statistic	Purpose	IV	DV	Relationship
Chi-square goodness of fit (x^2)	to test the predicted value of a proportion for a population.	-	Nominal	-
Chi-square test of independence (x^2)	to test the difference in proportions in 2+ independent groups.	Nominal	Nominal	phi (2x2) Cramer's v
Fisher's exact test	to test the difference in proportions (2 x 2 table) when expected frequency is <5 for a cell	Nominal	Nominal	phi
McNemar test (x^2)	to test the difference in proportions for 2 related groups (2 x 2 design)	Nominal	Nominal	phi
Cochran's test (Q)	to test the difference in proportions for 3+ related groups	Nominal	Nominal	-
Mann-Whitney U test (U)	to test the difference in the ranks of scores of 2 independent groups	Nominal	Ordinal	Glass rank biserial correlation
Kruskal Wallis (H)	to test the difference in the ranks of scores of 3+ independent groups	Nominal	Ordinal	Epsilon2
Wilcoxon signed ranked test (T or z)	to test the difference in the ranks of 2 related groups	Nominal	Ordinal	Matched pairs rank biserial correlation
Friedman test (x^2)	to test the difference in the ranks of scores of 3+ related groups	Nominal	Ordinal	Epsilon2
Spearman's rank (rs)	to test the existence of a relationship/correlation between two variables	Ordinal	Ordinal	(rs)
Kendall tau (t)	to test the existence of a relationship/correlation between two variables	Ordinal	Ordinal	tau

Note. Image taken from Creswell and Creswell (2018).

Limitations of Statistics

(1) Statistical law are accurate on average. Statistics are aggregates of measured facts, so a single observation or measure is not a statistic and should not be used to individualize care. Statistics deal with groups and aggregates only. The bigger the sample the more likely the data are likely to reflect the characteristics of the population being measured.

(2) Statistical methods are best applicable to quantitative data measured using a valid and reliable tool or instrument. If the instrument is not reliable, the data obtained using it may not valid and reliable.

(3) Statistics cannot be applied to heterogeneous data. If there are violations to the assumptions or the data is extremely different in the groups or characteristics of the sample, the results of the analyses may not be without error.

(4) If sufficient care is not exercised in collecting, analyzing and interpreting the data, statistical results might be misleading due to measurement error.

(5) Only a person who has an expert knowledge of statistics can handle statistical data efficiently to ensure that aspects of measurement error are identified. The statistical results should be examined differently than the clinical results as statistical significance is different than clinical significance.

(6) Some errors are possible in statistical decisions. In particular, inferential statistics involves certain errors. We do not know whether an error has been committed or not. We must assume the results will contain error until significant testing in other populations and settings shows the results can be reliably inferred.

WRITING UP YOUR STATISTICAL RESULTS

Identify the analysis technique.

In the results section, your goal is to report the results of the data analyses used to answer your project question. To do this, you need to identify your data analysis technique, report your test statistic, and provide some interpretation of the results. Each analysis you run should be related to your clinical question or PICOT. If you analyze data that is exploratory or outside your clinical question, you need to indicate this in the results.

Format test statistics.

Test statistics and *p* values ***should be rounded to two decimal places*** (If you are providing precise p-values for future use in meta-analyses, 3 decimal places is acceptable). ***All statistical symbols (sample statistics) that are not Greek letters should be italicized (M, SD, t, p, etc.).***

Indicate the direction of the significant difference.

When reporting a significant difference between two conditions, indicate the direction of this difference, i.e. which condition was more/less/higher/lower than the other condition(s). Assume that your audience has a professional knowledge of statistics. Do not explain how or why you used a certain test unless it is unusual (i.e., such as a non-parametric test like chi-square or Wilcoxon).

How to report *p* values.

Report the exact p value (this is the preferred option if you want to make your data convenient for individuals conducting a meta-analysis on the topic).

Example: $t(33) = 2.10$, $p = .03$.

If your exact p value is less than .001, it is conventional to state merely $p < .001$. If you report exact p values, state early in the results section the alpha level used as a significance criterion for your tests. For example: "We used an alpha level of .05 for all statistical tests."

If your results are in the predicted direction but are not significant, you can say your results were *marginally* significant. Example: Results indicated a marginally significant preference for pie ($M = 3.45$, $SD = 1.11$) over cake ($M = 3.00$, $SD = .80$), $t(5) = 1.25$, $p = .08$.

If your p-value is over .10, you can say your results revealed a non-significant trend in the predicted direction. Example: Results indicated a non-significant trending in the predicted

direction indicating a preference for pie ($M = 4.25$, $SD = 2.21$) over cake ($M = 3.25$, $SD = 2.60$), $t(5) = 1.75$, $p = .26$.

Descriptive Statistics

Mean and Standard Deviation are most clearly presented in parentheses:

The data in the Microsoft Excel was viewed for missing data, errors, and missing records: No

values were missing from the data. The descriptive variables of _____,

_____ and _____ for the sample (N = _____) were entered into an SPSS

version 28 database for descriptive analysis. The sample's descriptive variables were described

using mean, range, and standard deviation in narrative and displayed in a table (see Table 1). The

sociodemographic variables were reported in frequencies and percentages. Describe the findings

in you own words using these examples. The sample as a whole was relatively young ($M = \#$, SD

$= \#$). The average age of students was # years ($SD = \#$).

The frequencies for the sociodemographic variables of _____,

_____, _____and _____were tabled for the pre,

postimplementation and overall sample (see Table).   As noted in the table, the rates or

frequencies for _____were reported in counts and percentages for each variable.

Percentages are also most clearly displayed in parentheses with no decimal places. Describe what

is seen in the table.

Percentages are also most clearly displayed in parentheses with no decimal places:

☐ Nearly half (49%) of the sample was married.

Frequencies or rates are reported including the range, mode, or median.

***t*-tests**

There are several different designs that utilize a t-test for the statistical inference testing. The differences between one-sample t-tests, related measures t-tests, and independent samples t tests are clear to the knowledgeable reader so eliminate any elaboration of which type of t-test has been used. Additionally, the descriptive statistics provided will identify which variation was

employed. It is important to note that we assume that all *p* values represent two-tailed tests unless otherwise noted and that independent samples t-tests use the pooled variance approach (based on an equal variances assumption) unless otherwise noted:

☐ There was a significant effect for gender, $t(54) = 5.43$, $p < .001$, with men receiving higher scores than women.

☐ Results indicate a significant preference for pie ($M = 3.45$, $SD = 1.11$) over cake ($M = 3.00$, $SD = .80$), $t(15) = 4.00$, $p = .001$.

☐ The 36 study participants had a mean age of 27.4 ($SD = 12.6$) were significantly older than the university norm of 21.2 years, $t(35) = 2.95$, $p = .01$.

☐ Students taking statistics courses in psychology at the University of Washington reported studying more hours for tests ($M = 121$, $SD = 14.2$) than did UW college students in general, $t(33) = 2.10$, $p = .034$.

☐ The 25 participants had an average difference from pre-test to post-test anxiety scores of -4.8 ($SD = 5.5$), indicating the anxiety treatment resulted in a significant decrease in anxiety levels, $t(24) = -4.36$, $p = .005$ (one-tailed).

☐ The 36 participants in the treatment group ($M = 14.8$, $SD = 2.0$) and the 25 participants in the control group ($M = 16.6$, $SD = 2.5$), demonstrated a significance difference in performance ($t[59] = -3.12$, $p = .01$); as expected, the visual priming treatment inhibited performance on the phoneme recognition task.

☐ UW students taking statistics courses in Psychology had higher IQ scores ($M = 121$, $SD = 14.2$) than did those taking statistics courses in Statistics ($M = 117$, $SD = 10.3$), $t(44) =$

1.23, $p = .09$.

☐ Over a two-day period, participants drank significantly fewer drinks in the experimental group ($M = 0.667$, $SD = 1.15$) than did those in the wait-list control group ($M = 8.00$, $SD = 2.00$), $t(4) = -5.51$, $p = .005$.

ANOVA and post hoc tests

ANOVAs are reported like the t test, but there are two degrees-of-freedom numbers to report. First report the between-groups degrees of freedom, then report the within-groups degrees of freedom (separated by a comma). After that report the F statistic (rounded off to two decimal places) and the significance level.

One-way ANOVA:

☐ The 12 participants in the high dosage group had an average reaction time of 12.3 seconds ($SD = 4.1$); the 9 participants in the moderate dosage group had an average reaction time of 7.4 seconds ($SD = 2.3$), and the 8 participants in the control group had a mean of 6.6 ($SD = 3.1$). The effect of dosage, therefore, was significant, $F(2,26) = 8.76$,

p=.012.

☐ An one way analysis of variance showed that the effect of noise was significant, $F(3,27) = 5.94$, $p = .007$. Post hoc analyses using the Scheffé post hoc criterion for significance indicated that the average number of errors was significantly lower in the white noise condition ($M = 12.4$, $SD = 2.26$) than in the other two noise conditions (traffic and industrial) combined ($M = 13.62$, $SD = 5.56$), $F(3, 27) = 7.77$, $p = .042$.

☐ Tests of the four a priori hypotheses were conducted using Bonferroni adjusted alpha levels of .0125 per test (.05/4). Results indicated that the average number of errors was significantly lower in the silence condition ($M = 8.11$, $SD = 4.32$) than were those in both the white noise condition ($M = 12.4$, $SD = 2.26$), $F(1, 27) = 8.90$, $p =.011$ and in the industrial noise condition ($M = 15.28$, $SD = 3.30$), $F(1, 27) = 10.22$, $p = .007$. The pairwise comparison of the traffic noise condition with the silence condition was nonsignificant. The average number of errors in all noise conditions combined ($M = 15.2$, $SD = 6.32$) was significantly higher than those in the silence condition ($M = 8.11$, $SD = 3.30$), $F(1, 27) = 8.66$, $p = .009$.

Multiple Factor (Independent Variable) ANOVA (Requires more than two groups or data or more than two outcomes)

☐ There was a significant main effect for treatment, $F(1, 145) = 5.43$, $p < .01$, and a significant interaction, $F(2, 145) = 3.13$, $p < .05$.

☐ The cell sizes, means, and standard deviations for the 3x4 factorial design are presented in Table 1. The main effect of Dosage was marginally significant ($F[2,17] = 3.23$, $p = .067$), as was the main effect of diagnosis category, $F(3,17) = 2.87$, $p = .097$. The interaction of dosage and diagnosis, however, has significant, $F(6,17) = 14.2$, $p = .0005$.

☐ Attitude change scores were subjected to a two-way analysis of variance having two levels of message discrepancy (small, large) and two levels of source expertise (high, low). All effects were statistically significant at the .05 significance level. The main effect of message discrepancy yielded an F ratio of $F(1, 24) = 44.4$, $p < .001$, indicating that the mean change score was significantly greater for large-discrepancy messages ($M = 4.78$, $SD = 1.99$) than for small-discrepancy messages ($M = 2.17$, $SD = 1.25$). The main effect of source expertise yielded an F ratio of $F(1, 24) = 25.4$, $p < .01$, indicating that the mean change score was significantly higher in the high-expertise message source ($M = 5.49$, $SD = 2.25$) than in the low-expertise message source ($M = 0.88$, $SD = 1.21$). The interaction effect was non-significant, $F(1, 24) = 1.22$, $p > .05$.

☐ A two-way analysis of variance yielded a main effect for the diner's gender, $F(1,108) =$

3.93, $p < .05$, such that the average tip was significantly higher for men ($M = 15.3\%$, SD

$= 4.44$) than for women ($M = 12.6\%$, $SD = 6.18$). The main effect of touch was nonsignificant, $F(1, 108) = 2.24$, $p > .05$. However, the interaction effect was significant,

$F(1, 108) = 5.55$, $p < .05$, indicating that the gender effect was greater in the touch

condition than in the non-touch condition.

Chi Square

Chi-Square statistics are reported with degrees of freedom and sample size in parentheses, the Pearson chi-square value (rounded to two decimal places), and the significance level:

☐ The percentage of participants that were married did not differ by gender, $X^2(1, N = 90) = 0.89$, $p > .05$.

☐ The sample included 30 respondents who had never married, 54 who were married, 26

who reported being separated or divorced, and 16 who were widowed. These frequencies

were significantly different, $X^2 (3, N = 126) = 10.1$, $p = .017$.

☐ As can be seen by the frequencies cross tabulated in Table xx, there is a significant

relationship between marital status and depression, $X^2 (3, N = 126) = 24.7$, $p < .001$.

☐ The relation between these variables was significant, $X^2 (2, N = 170) = 14.14$, $p < .01$.

Catholic teens were less likely to show an interest in attending college than were

Protestant teens.

☐ Preference for the three sodas was not equally distributed in the population, $X^2 (2, N =$

$55) = 4.53$, $p = .05$.

Correlations

Correlations are reported with the degrees of freedom (which is N-2) in parentheses and the significance level:

☐ The two variables were strongly correlated, $r(55) = .49$, $p < .01$.

Tables

Add a table or figure.

Adding a table of figure can be helpful to the reader. See the current APA Publication manual for examples. In reporting the results of statistical tests, report the descriptive statistics, such as

20

means and standard deviations, as well as the test statistic, degrees of freedom, obtained value of the test, and the probability of the result occurring by chance (p value).

•APA style tables do not contain any vertical lines

•There are no periods used after the table number or title.

•When using columns with decimal numbers, make the decimal points line up.

•Use MS Word tables to create tables

American Psychological Association [APA]. (2019). *Publication manual of the American Psychological Association* (7th ed.). Washington, DC: Author.

Table 1

Sample Demographic Characteristics Table

Baseline characteristic	Preimplementation n =			Postimplementation n =		
	M	SD	Range	*M*	SD	Range
Variable						
Variable						

Table 1

Sociodemographic Characteristics of Participants at Baseline

Baseline characteristic	Guided self-help		Unguided self-help		Wait-list control		Full sample	
	n	%	*N*	%	*n*	%	*n*	%
Gender								
Female	25	50	20	40	23	46	68	45.3
Male	25	50	30	60	27	54	82	54.7
Marital status								
Single	13	26	11	22	17	34	41	27.3
Married/partnered	35	70	38	76	28	56	101	67.3
Divorced/widowed	1	2	1	2	4	8	6	4.0
Other	1	1	0	0	1	2	2	1.3
Children [a]	26	52	26	52	22	44	74	49.3
Cohabitating	37	74	36	72	26	52	99	66.0
Highest educational level								
Middle school	0	0	1	2	1	2	2	1.3
High school/some college	22	44	17	34	13	26	52	34.7
University or postgraduate degree	27	54	30	60	32	64	89	59.3
Employment								
Unemployed	3	6	5	10	2	4	10	6.7
Student	8	16	7	14	3	6	18	12.0
Employed	30	60	29	58	40	80	99	66.0
Self-employed	9	18	7	14	5	10	21	14.0
Retired	0	0	2	4	0	0	2	1.3

22

Previous psychological treatment [a]	17	34	18	36	24	48	59	39.3
Previous psychotropic medication [a]	6	12	13	26	11	22	30	20.0

Note. $N = 150$ ($n = 50$ for each condition). Participants were on average 39.5 years old ($SD =$

10.1), and participant age did not differ by condition.

[a] Reflects the number and percentage of participants answering "yes" to this question.

Sample Results of Several *t* Tests Table
Table 2

Results of Curve-Fitting Analysis Examining the Time Course of Fixations to the Target

Logistic parameter	9-year-olds		16-year-olds		$t(40)$	p	Cohen's
	M	*SD*	*M*	*SD*			*d*
Maximum asymptote, proportion	.843	.135	.877	.082	0.951	.347	0.302
Crossover, in ms	759	87	694	42	2.877	.006	0.840
Slope, as change in proportion per ms	.001	.0002	.002	.0002	2.635	.012	2.078

Note. For each subject, the logistic function was fit to target fixations separately. The maximum asymptote is the asymptotic degree of looking at the end of the time course of fixations. The crossover point is the point in time the function crosses the midway point between peak and baseline. The slope represents the rate of change in the function measured at the crossover. Mean parameter values for each of the analyses are shown for the 9-year-olds ($n = 24$) and 16-year-olds ($n = 18$), as well as the results of *t* tests (assuming unequal variance) comparing the parameter estimates between the two ages.

Sample Correlation Table
Table 1

Descriptive Statistics and Correlations for Study Variables

Variable	*n*	*M*	*SD*	1	2	3	4	5	6	7
1. Internal–external status [a]	3,697	0.43	0.49	—						
2. Manager job performance	2,134	3.14	0.62	−.08**	—					
3. Starting salary [b]	3,697	1.01	0.27	.45**	−.01	—				
4. Subsequent promotion	3,697	0.33	0.47	.08**	−.07**	.04*	—			
5. Organizational tenure	3,697	6.45	6.62	−.29**	.09**	.01	.09**	—		
6. Unit service performance [c]	3,505	85.00	6.98	−.25**	−.39**	.24**	.08**	.01	—	
7. Unit financial performance [c]	694	42.61	5.86	.00	−.03	.12*	−.07	−.02	.16**	—

[a] 0 = internal hires and 1 = external hires.

[b] A linear transformation was performed on the starting salary values to maintain pay practice confidentiality. The standard deviation (0.27) can be interpreted as 27% of the average starting salary for all managers. Thus, ±1 *SD* includes a range of starting salaries from 73% (i.e., 1.00 − 0.27) to 127% (i.e., 1.00 + 0.27) of the average starting salaries for all managers.

[c] Values reflect the average across 3 years of data.

*$p < .05$. **$p < .01$.

Table 1

Means, Standard Deviations, and One-Way Analyses of Variance in Psychological and Social

Resources and Cognitive Appraisals

Measure	Urban		Rural		$F(1, 294)$	η^2
	M	SD	M	SD		
Self-esteem	2.91	0.49	3.35	0.35	68.87[***]	.19
Social support	4.22	1.50	5.56	1.20	62.60[***]	.17
Cognitive appraisals						
Threat	2.78	0.87	1.99	0.88	56.35[***]	.20
Challenge	2.48	0.88	2.83	1.20	7.87[***]	.03
Self-efficacy	2.65	0.79	3.53	0.92	56.35[***]	.16

[***]$p < .001$.

Table 1

Results From a Factor Analysis of the Parental Care and Tenderness (PCAT) Questionnaire

PCAT item	Factor loading		
	1	2	3
Factor 1: Tenderness—Positive			
20. You make a baby laugh over and over again by making silly faces.	**.86**	.04	.01
22. A child blows you kisses to say goodbye.	**.85**	−.02	−.01
16. A newborn baby curls its hand around your finger.	**.84**	−.06	.00
19. You watch as a toddler takes their first step and tumbles gently back down.	**.77**	.05	−.07
25. You see a father tossing his giggling baby up into the air as a game.	**.70**	.10	−.03
Factor 2: Liking			
5. I think that kids are annoying (R)	−.01	**.95**	.06
8. I can't stand how children whine all the time (R)	−.12	**.83**	−.03
2. When I hear a child crying, my first thought is "shut up!" (R)	.04	**.72**	.01
11. I don't like to be around babies. (R)	.11	**.70**	−.01
14. If I could, I would hire a nanny to take care of my children. (R)	.08	**.58**	−.02
Factor 3: Protection			
7. I would hurt anyone who was a threat to a child.	−.13	−.02	**.95**
12. I would show no mercy to someone who was a danger to a child.	.00	−.05	**.74**
15. I would use any means necessary to protect a child, even if I had to hurt others.	.06	.08	**.72**
4. I would feel compelled to punish anyone who tried to harm a child.	.07	.03	**.68**
9. I would sooner go to bed hungry than let a child go without food.	.46	−.03	**.36**

Note. $N = 307$. The extraction method was principal axis factoring with an oblique (Promax with

Kaiser Normalization) rotation. Factor loadings above .30 are in bold. Reverse-scored items are

denoted with an (R). Adapted from "Individual Differences in Activation of the Parental Care

Motivational System: Assessment, Prediction, and Implications," by E. E. Buckels, A. T. Beall,

M. K. Hofer, E. Y. Lin, Z. Zhou, and M. Schaller, 2015, *Journal of Personality and Social*

Psychology, *108*(3), p. 501 (https://doi.org/10.1037/pspp0000023). Copyright 2015 by the

American Psychological Association.

Sample Regression Table
Table 3

Moderator Analysis: Types of Measurement and Study Year

Effect	Estimate	*SE*	95% CI		*p*
			LL	*UL*	
Fixed effects					
Intercept	.119	.040	.041	.198	.003
Creativity measurement [a]	.097	.028	.042	.153	.001
Academic achievement measurement [b]	−.039	.018	−.074	−.004	.03
Study year [c]	.0002	.001	−.001	.002	.76
Goal [d]	−.003	.029	−.060	.054	.91
Published [e]	.054	.030	−.005	.114	.07
Random effects					
Within-study variance	.009	.001	.008	.011	<.001
Between-study variance	.018	.003	.012	.023	<.001

Note. Number of studies = 120, number of effects = 782, total *N* = 52,578. CI = confidence

interval; *LL* = lower limit; *UL* = upper limit.

[a] 0 = self-report, 1 = test. [b] 0 = test, 1 = grade point average. [c] Study year was grand centered. [d] 0

= other, 1 = yes. [e] 0 = no, 1 = yes.

Sample Qualitative Table With Variable Descriptions

Table 2

Master Narrative Voices: Struggle and Success and Emancipation

Discourse and dimension	Example quote
Struggle and success [a]	
Self-actualization as member of a larger gay community is the end goal of healthy sexual identity development, or "coming out"	"My path of gayness ... going from denial to saying, well this is it, and then the process of coming out, and the process of just sort of, looking around and seeing, well where do I stand in the world, and sort of having, uh, political feelings." (Carl, age 50)
Maintaining healthy sexual identity entails vigilance against internalization of societal discrimination	"When I'm like thinking of criticisms of more mainstream gay culture, I try to ... make sure it's coming from an appropriate place and not like a place of self-loathing." (Patrick, age 20)
Emancipation [b]	
Open exploration of an individually fluid sexual self is the goal of healthy sexual identity development	"[For heterosexuals] the man penetrates the female, whereas with gay people, I feel like there is this potential for really playing around with that model a lot, you know, and just experimenting and exploring." (Orion, age 31)
Questioning discrete, monolithic categories of sexual identity	"LGBTQI, you know, and added on so many letters. Um, and it does start to raise the question about what the terms mean and whether ... any term can adequately be descriptive." (Bill, age 50)

[a] The struggle and success master narrative states that same-sex desire/behavior is a natural if relatively uncommon developmental variant distinguishable from heterosexuality. Healthy sexual development entails "coming out" as well as joining a larger gay community in a shared struggle to overcome societal discrimination and be socially recognized as normal.

[b] The emancipation master narrative states that discrete, monolithic, and mutually exclusive categories of homosexuality and heterosexuality are social constructions, conceptually suspect in their ability to fully capture the idiosyncrasies of sexual subjectivities, desires, and behaviors. This circumscription of sexual self within culturally contingent and hegemonic sexual identity categories must be resisted.

Sample Mixed Methods Table
Table 3

Integrated Results Matrix for the Effect of Topic Familiarity on Reliance on Author Expertise

Quantitative results	Qualitative results	Example quote
When the topic was more familiar (climate change) and cards were more relevant, participants placed less value on author expertise.	When an assertion was considered to be more familiar and considered to be general knowledge, participants perceived less need to rely on author expertise.	Participant 144: "I feel that I know more about climate and there are several things on the climate cards that are obvious, and that if I sort of know it already, then the source is not so critical ... whereas with nuclear energy, I don't know so much so then I'm maybe more interested in who says what."
When the topic was less familiar (nuclear power) and cards were more relevant, participants placed more value on authors with higher expertise.	When an assertion was considered to be less familiar and not general knowledge, participants perceived more need to rely on author expertise.	Participant 3: "[Nuclear power], which I know much, much less about, I would back up my arguments more with what I trust from the professors."

Note. We integrated quantitative data (whether students selected a card about nuclear power or about climate change) and qualitative data (interviews with students) to provide a more comprehensive description of students' card selections between the two topics.

References

American Psychological Association [APA]. (2019). *Publication manual of the American Psychological Association* (7th ed.). Washington, DC: Author. ISBN: ISBN-13: 978-1433832178, ISBN-10: 1433832178

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences.* New York, NY: Routledge Academic.

Doherty, C. D., & Skalsky, K. (2021). *Statistics and research design for the DNP Project*. Destech Publications ISBN 160593849.

Grove, S., & Cipher, D. (2019). *Statistics for nursing research.* (3rd e.d.). Elsevier.

McHugh M. L. (2013). The chi-square test of independence. *Biochemia medica*, *23*(2), 143–149. https://doi.org/10.11613/bm.2013.018

Melnyk, B. M., & Fineout-Overholt, E., (2018). *Evidence-based practice in nursing and healthcare: A guide to best practice*. 4th Edition. Lippincott Williams & Wilkins. ISBN-13: 9781496384539

Polit, D. F., & Beck, C. T. (2021). Nursing research: generating and assessing evidence for nursing practice. Eleventh edition. Philadelphia, Wolters Kluwer.

Zaccagnini, M. E., & White, K. (2022). The doctor of nursing practice essentials (4th ed.). Jones and Bartlett.