

Reducing Error: Averaging Data to Determine Factor Structure of the QMPR

Shea Gibbons¹, Robert Bubb², Bruce Brown³

¹ Psychology, Brigham Young University

² Psychology, Auburn University, 206 Thach Hall, Auburn University, AL 36849

³ Psychology, Brigham Young University, 1098 SWKT, Provo, UT, 84604

Abstract

Factor loading instability is common in single administration human subject data. Factor loading instability results in poor interpretation of the factor pattern. However, Monte Carlo simulations have shown that averaging data across multiple administrations reduces unexplained error, increases the amount of explained variance, and stabilizes factor patterns. In the present study, the Questionnaire for the Measurement of Psychological Reactance (QMPR) was administered to participants multiple times. The data were averaged and then analyzed using principle components factor analysis with varimax rotation. The study results demonstrated a significant increase in explained variance and factor pattern stability and improved interpretability.

Key Words: Factor Analysis, Reliability, Averaging Data, Human Subject Data, QMPR

1. Introduction

Self-report behavior in psychological research is inconsistent and unstable (Epstein, 1979). One way to increase consistency and stability of self-report behavior is to average multiple observations for each participant. As responses are averaged, the error associated with self-reported behavior will decrease (Cronbach, 1970). Using an average to represent a sample is a fundamental tenet of statistics. However, averaging data across replications to represent an individual has not been an accepted practice. With the exception of repeated measures designs, psychological research tends to utilize a single administration design. In 1980, Seymour Epstein questioned the single administration design (795), “Every beginning psychology student knows that findings from a single individual should not be generalized to other people. By what logic does one then assume that findings from a single stimulus situation or from a restricted range of stimuli can yield generalizations that hold for other stimuli and stimulus situations?” Epstein cites a number of studies in which averaging data reduced measurement error that inevitably accompanies behavioral research (see Epstein, 1979). As a simple solution to measurement error, averaging produces reliable results.

Factor analysis in psychological research allows researchers to define variables more precisely, show relationships between variables, and define phenomena to be studied more specifically (Comfrey, 1992). In order to develop validated instruments that have precise variables that are related and well defined, it is critical to reduce the amount of measurement error in data. Bubb, Brown, and Pedersen (2007) demonstrated with Monte Carlo simulations that measurement error in factor analytic research can be reduced by averaging data across administrations (see figures 1 and 2). Averaging data produced greater consistency and stability in factor loadings and increased the amount of explained variance (see figures 1 and 2). Consistent and stable latent structures are necessary to validate psychological diagnosis instruments.

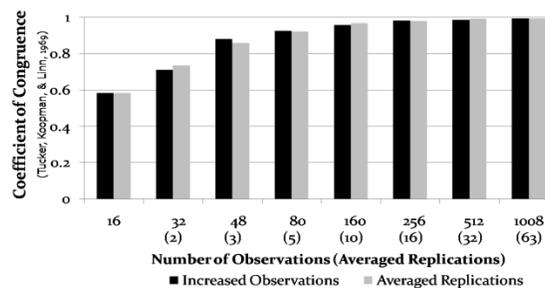


Figure 1. The comparison of values for the coefficient of congruence from increasing the sample size vs. averaging observations (given in parentheses)

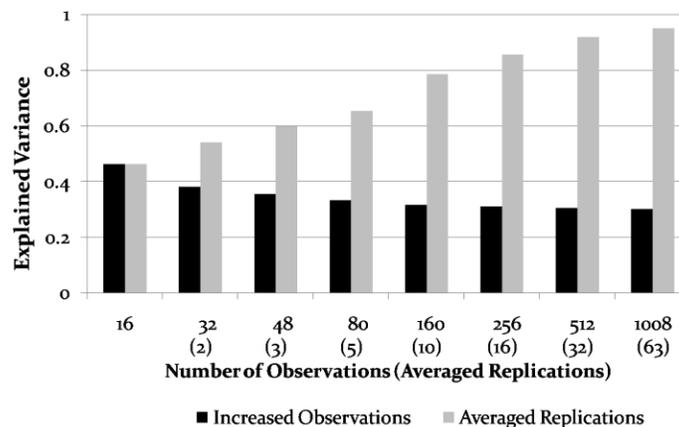


Figure 2. The comparison of values for the explained variance from increasing the sample size vs. averaging observations (given in parentheses)

In the present study, the researchers attempted to verify the Monte Carlo simulation findings of Bubb et al. (2007) using human subject data rather than simulated data. In the Bubb et al. (2007) study, errorless data was artificially created and combined with random data. Random data was used to simulate measurement error; however, not all measurement error is random. Non-random error can be classified as explained error or biases. Biases that may confound results might include distribution errors such as halo, leniency, and strictness and observation effects such as recency, primacy, and consistency (Aamodt, 2007). Biases reduce the validity of findings and if biases are consistent across self-reports, then averaging data may have less of an effect on increasing factor pattern stability and structural validity. To determine the extent to which averaging data across administrations reduces random error, as opposed to explained error, data using human subjects is necessary.

1.1 Questionnaire for the Measurement of Psychological Reactance (QMPR)

In order to factor analyze human subject data averaged across administrations, a questionnaire with low factor pattern stability is required. Low factor pattern stability implies a high amount of measurement error. The questionnaire selected for this study, the Questionnaire for the Measurement of Psychological Reactance (QMPR), attempts to quantify the latent construct of psychological reactance. The QMPR was created by Merz (1983). Reactance is the concept that people defy attempts by authority to restrict personal freedom of thoughts or behaviors (Brehm, 1966). It is hypothesized that reactance is related to aggression, altruism, depression, mood change, frustration, social power, self-awareness, locus of control, and compliance (Merz, 1983; Brehm, 1966; Swat, Ickes, & Morgenthaler, 1978; Moyer, 1978). Despite the apparent benefits of understanding more about reactance, few research studies have analyzed the components of reactance (Tucker & Byers, 1987). Merz' questionnaire seeks to identify those components.

Merz developed the QMPR in German by drafting 32 items and submitting them to four professional psychologists for evaluation (Tucker & Byers, 1987). Any items that two or more of the psychologists deemed unfit were excluded. He then administered the revised 26 item questionnaire to 152 high school and university students. Based on the results of that administration, 8 more questions were eliminated. The result was an 18-item questionnaire. The questionnaire was measured on a 6-point Likert scale. Merz used a principle components factor analysis with varimax rotation to analyze the data. A 4-factor solution was interpreted from the factor analysis that explained 53% of the variance. However, Merz did not report the criteria for extracting the factors, leaving many researchers wondering as to how he arrived at a 4-factor solution (Tucker & Byers, 1987; Donnell, et.al., 2001).

In an attempt to validate Merz' measure, Tucker and Byers (1987) replicated the study with an English version of the QMPR and an American sample of 218 college students. They used a 5-point Likert scale. The 5-point scale deviated from the 6-point scale used by Merz, drawing criticism from Donnell, Thomas, and Buboltz (2001). The analysis of Tucker and Byers (1987) indicated that a 2-factor solution, explaining 21% of the variance, was most appropriate. They conceded that "translation loss" may have occurred, but also stated that the low factor loadings and high amounts of error indicate that the QMPR is psychometrically unstable.

In 1989, Hong and Ostini sought to validate Merz' QMPR with an Australian sample of 379 tertiary students. They retained the QMPR as translated in the Tucker and Byers study, but instead used a 4-point Likert scale. Hong and Ostini interpreted a 4-factor solution that accounted for 44.1% of the variance. They also conclude that the QMPR is psychometrically unstable and suggest revisions to improve the readability of the instrument in English.

In 2001, Donnell et al. researched the psychometric properties of the QMPR with an American sample of 898. By using a larger sample size than the previous studies, Donnell et al. hoped to finally determine the psychometric properties of the QMPR. Obtaining a larger sample size increases correlation coefficient stability and results in greater stability in the factor pattern (Cattell, 1952). The researchers re-translated the QMPR from Merz, resulting in a more readable and understandable version. Donnell et al. also used a 6-point Likert scale to model the original Merz study. The factor analytic interpretation by Donnell et al. yielded a 3-factor solution that explained 38.3% of the variance. Donnell et al. also concluded that the QMPR was psychometrically unsatisfactory.

As shown in Table 1, the one-time administrations by four different studies produced divergent factor solutions and explained varying amounts of the common variance. Each of the preceding validation studies of the QMPR failed to produce a stable factor pattern, making the QMPR a suitable questionnaire to which the method of averaging individual subject data over multiple administrations can be applied. It is anticipated that averaging will provide a stable factor pattern as measured by the coefficient of congruence (Tucker, Koopman, & Linn, 1969) and an increased amount of explained variance than in the previously published validation studies. The highest reported variance was 53% (Merz, 1983). Although factor pattern stability and increased variance will not validate the instrument, they will lead to a clearer factor pattern. Items in the factor pattern not clearly loading onto a single factor or loading onto multiple factors will be more easily identified as poor items; therefore subject to adjustment or removal. Clearer interpretations of factor patterns lead to a better process for developing instruments and increases the quality of validated instruments.

Table 1
Results of the previous QMPR studies

	Merz (1983)	Tucker & Byers (1987)	Hong & Ostini (1989)	Donnell et al. (2001)
Extraction	Principal components	Principal components	Principal axis	Principal components and maximum likelihood
Rotation	Varimax	Direct Quartimin	Varimax	Varimax & Direct oblimin
Factors	4	2	4	3
% variance	53	21	44	38.3
Sample size	152	218	379	898
Likert scale	6 points	5 points	4 points	6 points

2. Method

2.1 Subjects

The participants for this study were a sample of 143 male and female psychology major and minor students in various psychology courses. Students were offered extra-credit to participate in the study.

The 18-question (scaled using a 6-point Likert type scale where 1 = does not apply at all, 6 = always applies) QMPR as translated in the Donnell et al. (2001) study was administered via email once every two days for 42 days. Respondents clicked on a survey link provided in the email. The link led to a Qualtrics online survey. Each respondent completed the measure 10 times. Emails were discontinued once the respondent completed 10 administrations. Respondent responses were automatically recorded.

2.2 Procedures

Once the data were collected, each subject's data were averaged over multiple administrations of 2 through 10. For each averaged administration (except for 10 since there is only one possible averaged result), 10 different sets were obtained through random selection. For example, of the 45 possible combinations of averages of 2 administrations, 10 were chosen by random selection. The selection of 10 combinations (sets) was conducted for each of the averages of two through nine. Additionally, the 10 single administration events were also selected as sets, as well as the average of all 10 administrations. There were a total of 91 sets. Each set was then factor analyzed using exploratory principal components factor analysis with varimax rotation in SPSS 15.0. Principle components and varimax rotation were deliberately selected to allow for comparison with the analysis conducted by Merz (1983). The coefficient of

congruence was used to compare factor patterns within each set. The coefficient of congruence is the correlation between two matrices of the same size (Tucker, Koopman, & Linn, 1969). Coefficients of congruence and the amount of explained variance were averaged across sets within each averaged administration 1 through 9¹. Once factor pattern stability, as measured by the coefficient of congruence, and explained variance were calculated for each set, then factor pattern stability and the amount of explained variance were compared among averaged administrations. Multivariate analysis of variance (MANOVA) statistics were obtained to compare the averaged administrations where the levels of the independent variable were the averaged administrations 1 through 9 and the dependent variables were the coefficient of congruence and the amount of variance explained. It was hypothesized that the coefficients of congruence and amount of explained variance means would increase and the standard deviations would decrease as the number of averaged administrations increased from 1 to 9.

3. Results

As hypothesized, the coefficient of congruence and the amount of explained variance increased and the standard deviations decreased as the number of averaged administrations increased from 1 to 9 (see figures 3 and 4 and Table 2). MANOVA statistics, with the levels of the independent variable as the averaged administrations 1 through 9 and the dependent variables as the coefficient of congruence and amount of explained variance, demonstrated a significant effect. All four multivariate tests were significant at $p < .0001$ with Wilk's Lambda as .137, Pillai's Trace as .968, Hotelling-Lawley Trace as 5.518, and Roy's Greatest Root as 5.375. The individual analysis of variance (ANOVA) result for the coefficient of congruence in the data was statistically significant with $F = 16.502$, $p < .0001$. The coefficient of congruence increased as the number of administrations averaged increased (single administration $m = .882$, $sd = .086$; 9 administrations averaged $m = .999$, $sd = .0002$) accounting for 58% of the variance in the data. The individual ANOVA results for the amount of variance explained was also statistically significant with $F = 22.555$, $p < .0001$. The amount of explained variance with four factors increased as the number of administrations averaged increased (single administration $m = 57.81$, $sd = 4.27$; 9 administrations averaged $m = 66.16$, $sd = .309$) accounting for 66% of the variance in the data. Differences in the coefficient of congruence are visualized in factor patterns in figure 5 for single administrations, five administrations averaged together, and nine administrations averaged together.

Table 2
Results for each set of averaged administrations

Number of Averages	Variance explained		Coefficient of Congruence	
	Mean	Standard Deviation	Mean	Standard Deviation
1	57.81	4.27	0.8823	0.0865
2	60.69	2.67	0.9779	0.0094
3	64.20	1.24	0.9887	0.0050
4	63.92	1.60	0.9886	0.0053
5	64.93	1.00	0.9931	0.0023
6	65.56	0.77	0.9967	0.0017
7	65.84	0.45	0.9979	0.0008
8	65.67	0.52	0.9984	0.0008
9	66.16	0.31	0.9995	0.0002

¹ Single administrations are referred to averages of 1 in this study, i.e. averaged with self.

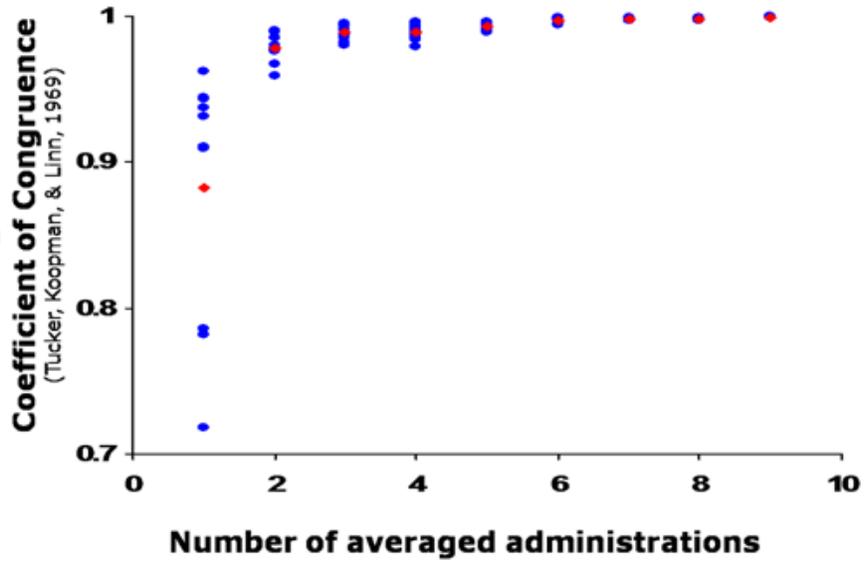


Figure 3. The effect of averaging on the coefficient of congruence (Tucker, Koopman, & Linn, 1969).

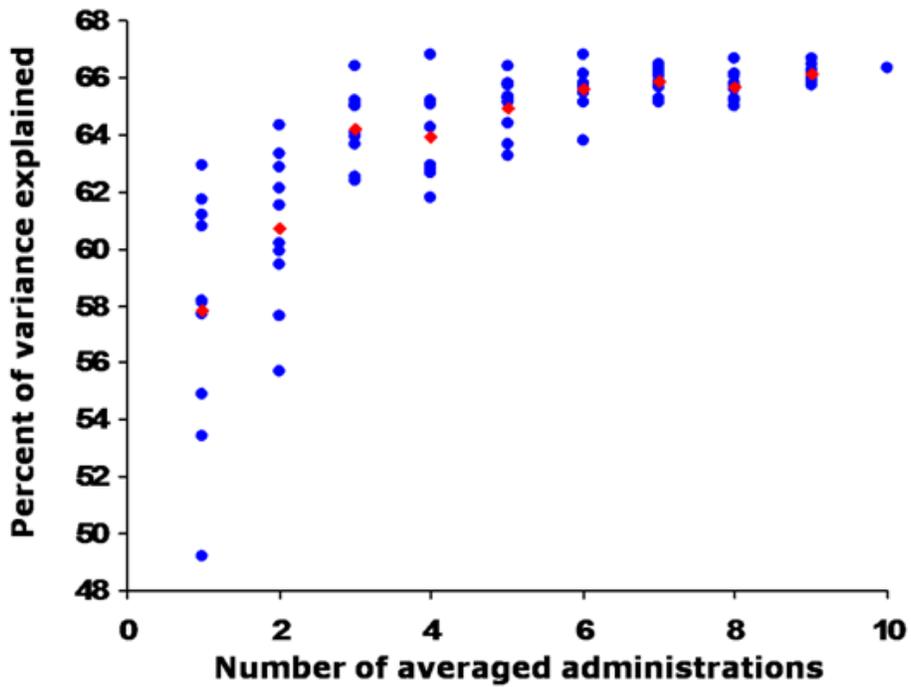


Figure 4. The effect of averaging on the explained variance.

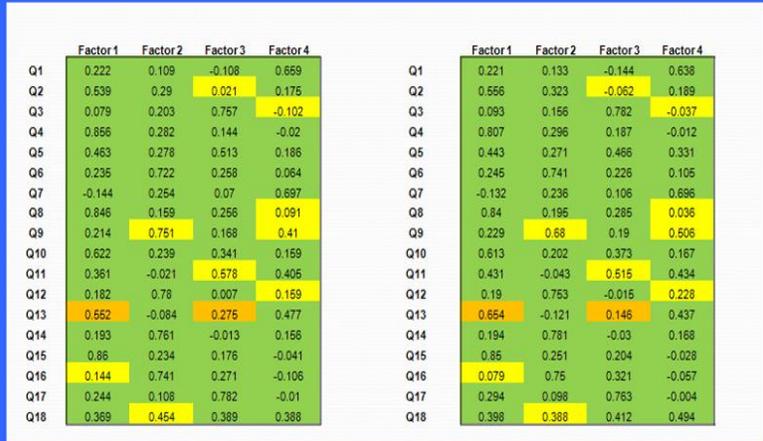
One Administration

Coefficient of Congruence = 0.882



Averages of Five Administrations

Coefficient of Congruence = 0.993



Averages of Nine Administrations

Coefficient of Congruence = 0.9995

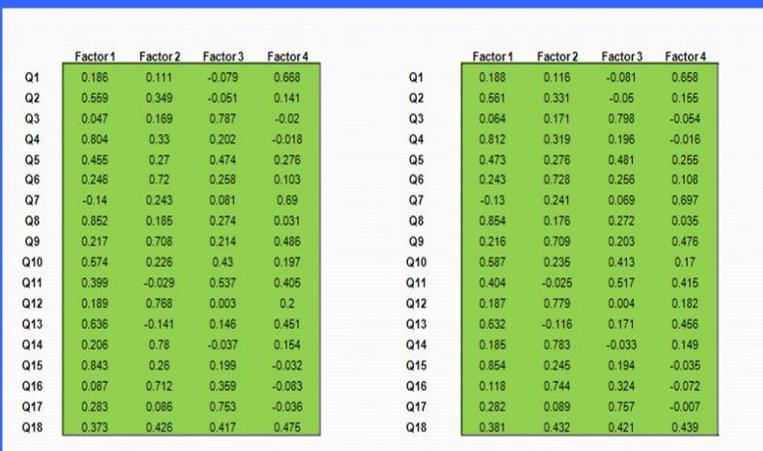


Figure 5. The factor patterns for the averages of one, five, and nine administrations with the corresponding values for the coefficient of congruence. Red marks indicate the differences for factor loadings between the two scores are greater than .15. Orange marks indicate the differences are from .10-.15. Yellow marks indicate that the differences are from .05 to .10. Green marks indicate the differences are .05 or less.

4. Discussion

The results supported the hypotheses. Factor loadings became more stable as averaging increased. The amount of explained variance also increased as more administrations were averaged. In comparison to the simulation study by Bubb et al. (2007) study, the coefficient of congruence was slightly higher in the present study ($K = .9995$ to $.9676$), indicating a more stable factor pattern; however, the extent to which the amount of explained variance improved due to averaging was not as drastic. The amount of explained variance of 66% was higher than the highest reported variance of 53% reported in the previous QMPR studies (Merz, 1983). However, it did not reach the level of 78% as reported in the Bubb et al. (2007) study over ten averaged administrations. Part of the discrepancy can be attributed to the factor analytic model. The Bubb et al. study used a 3-factor, 24 variable model, averaged over 10 administration; whereas the present research used a 4 factor, 18 variable model, averaged over 9 administrations. Bubb et al. (2007) reported that increasing the number of factors, decreasing the number of variables, and decreasing the number of averaged administrations led to a decrease in the amount of explained variance. It is possible given these discrepancies that the difference between the simulation data and the human subject data is less significant than reported in the present study.

Any remaining discrepancy in the amount of unexplained variance may be due to non-random error or biases common in human subject data, such as distribution errors, observation bias, unrepresentativeness of the factors, and poor item construction. Despite the discrepancy in the amount of variance explained between the Bubb et al. study and the current study, factor pattern stability was obtained by averaging 9 administrations. A coefficient of congruence of $.99$ was obtained after only 5 administrations. Averaging as few as 5 times may be sufficient to produce factor pattern stability, and thus making it easier to determine the appropriateness of items in an instrument. Averaging individual subject data across multiple administrations may lead to more cost-efficient methods of validating instruments. For example, the factor pattern averaged over 9 times allows for the easy identification of items that load onto multiple factors. Identifying items that are not correlated with a single construct may help researchers to decide whether to reword or eliminate items. Averaging across multiple administrations for the same person led to clear, stable, and easily interpretable factor loadings and increased the amount of explained variance. The largest decrease in error occurred after two administrations were averaged. A test averaged over two administrations would lead to an improvement in interpretation, with a further increase in the amount of explained variance and factor pattern stability with a test averaged over five administrations.

Additionally, these results paired with the Bubb et al. (2007) results show that the recommendation to produce stability by simply maximizing sample size may be misguided. Increasing the sample size does produce factor pattern stability, but the opposite effect is found concerning the amount of explained variance, which decreases until it levels out as the sample size increases (see figures 1 and 2). Averaged administration results also show a more accurate picture of the individuals in the study than a single administration, large-N study. Random error is decreased by averaging.

4.1 Additional Implications

The number of subjects in this study was acceptable to many divergent guidelines governing the ideal number of subjects for factor analytic studies (see Bubb et al., 2007). Even with an acceptable number of subjects in this study, the factor pattern comparisons for the first administration was poor compared with the factor patterns that resulted from averaging (see figure 5). Poor comparisons among single administrations have startling implications for single administration factor analytic studies that are reported in many scientific journals: they may not offer a clear, consistent factor solution. An unclear factor solution may lead to confusion over which instrument items are reliably loading onto which factors. For example, many items in the factor pattern shown in figure 5 load onto two separate factors or there is a lower coefficient of congruence. Researchers who utilize a one-time administration may eliminate reliable items or accept poor items loading onto a factor due to error inherent in one-time administrations. Care should be taken in using only stable and validated instruments when conducting research.

4.2 Limitations

Evidence of a consistency bias occurs in the data. There was high test-retest reliability between administrations. This may be partially due to the validity of the instrument or to practice effects. However, informal exit interviews with respondents revealed a desire to maintain consistency between administrations. Maintaining consistency among administrations reduced variability in the level of reactance experienced at each administration. In future studies, participants should be instructed to focus on their current level of reactance rather than trying to remain consistent with previous administrations of the QMPR. Further research using averaged administration data should increase the amount of time between administrations or combine the QMPR with other studies to reduce practice

effects. Additionally, since the benefits of averaging level out quickly, three to five administrations may be sufficient to maximize the benefits and reduce the practice effects attributed to later administrations.

Furthermore, while self-report surveys report an intuitive averaging by the respondent, self-report surveys do not necessarily correlate with behavior. Epstein (1983) argues that when a respondent completes an item on a survey, the respondent may recall salient and specific instances that may misrepresent the average behavior of the instrument item. On the other hand, low correlations between attitudes and behavior also complicate the conclusions that may be drawn from self-report research.

Another limitation to the present study is the large amount of shared data in the higher numbered averaged administrations. For example, one of the averages of seven consisted of administrations 1, 2, 4, 5, 7, 8, and 10 and another consisted of administrations 2, 4, 5, 6, 7, 9, and 10. In this case, four of the seven sets that were averaged were identical; thereby increasing the coefficient of congruence since anything correlated with itself will equal 1. This is an unavoidable aspect of the study. However, the results in the lower numbered averages, where less shared data exists, present sufficient evidence to support the hypotheses. Averaging shared data had a higher incident rate in the averaged administrations of 6 or more, whereas averaged administrations of 5 or below were less likely to be randomly selected twice.

Acknowledgements

The author would like to sincerely thank both co-authors, Robert Bubb and Dr. Bruce Brown, both of whom have been instrumental in making the research happen. Robert Bubb in particular performed a special service in the mentoring provided.

References

- Aamodt, M.G. (2007). *Industrial and organizational psychology*. Belmont, CA: Thomson Wadsworth.
- Brehm, J. W. (1966). *A theory of psychological reactance*. San Diego, CA: Academic Press.
- Bubb, R. R., Brown, B. L., & Pedersen, D. M. (2007) Determining a minimum number of observations in factor analytic studies: Model Construct Error Effects in Obtaining Factor Loading Stability and Maintaining Structural Integrity. *American Statistical Association 2007 Proceedings*. Salt Lake City: American Statistical Association.
- Cattell, R. B. (1952). *Factor analysis: An introduction and manual for the psychologist and socialist scientist*. New York: Harper & Brothers.
- Child, D. (1990). *The essentials of factor analysis* (2nd ed.). London: Cassell Educational Limited.
- Comfrey, A. L., & Lee, H. B. (1992). *A first course in factor analysis*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.
- Cronbach, L. J. (1970). *Essentials of psychological testing* (3rd ed.). New York: Harper & Row.
- Donnell, A., Thomas, A., & Buboltz, W. (2001). Psychological reactance: Factor structure and internal consistency of the Questionnaire for the Measurement of Psychological Reactance. *The Journal of Social Psychology, 141*, 679–687.
- Epstein, S. (1979). The stability of behavior: I. On predicting most of the people much of the time. *Journal of Personality and Social Psychology, 37*, 1097-1126.
- Epstein, S. (1980). The stability of behavior: II. Implications for psychological research. *American Psychologist, 35*, 790-806.
- Epstein, S. (1983). Aggregation and beyond: Some basic issues on the prediction of behavior. *Journal of Personality, 51*, 360-392.
- Hong, S.-M., & Ostini, R. (1989). Further evaluation of Merz's psychological reactance scale. *Psychological Reports, 64*, 707–710.
- Merz, J. (1983). Fragebogen zur Messung der Psychologischen Reaktanz [Questionnaire for Measuring Psychological Reactance]. *Diagnostica, 29*, 75–82.
- Moyer, W. W. (1978). Effects of loss of freedom on subjects with internal or external locus of control. *Journal of Research in Personality, 12*, 253-261.
- Swart, C., Ickes, W., & Morgenthaler, E. S. (1978). The effect of objective self awareness on compliance in a reactance situation. *Social Behavior and Personality, 6*, 135-139.
- Tucker, L. R., Koopman, R. F., & Linn, R. L. (1969). Evaluation of factor analytic research procedures by means of simulated correlation matrices. *Psychometrika, 34*, 421-459.
- Tucker, R. K., & Byers, P. Y. (1987). Factorial validity of Merz's psychological reactance scale. *Psychological Reports, 61*, 811–815.