Included Analyses

- Descriptive Statistics
- Pearson Correlations for DP, EE, and PA
- Independent Samples t-Test for DP by Gender
- Mann-Whitney Rank Sum Test for DP by Gender
- Linear Regression with DP predicted by EE and PA

Results

Descriptive Statistics

Introduction

Summary statistics were calculated for each interval and ratio variable. Frequencies and percentages were calculated for each nominal variable.

Frequencies and Percentages

The most frequently observed category of Gender was Female (n = 8, 53.33%). Frequencies and percentages are presented in Table 1.

Table 1

Frequency Table for Nominal Variables

Variable	*n*	%
Gender		
Male	7	46.67

Female	8	53.33

Note. Due to rounding errors, percentages may not equal 100%.

Summary Statistics

The observations for DP had an average of 49.07 (SD = 32.22, SE_M = 8.32, Min = 12.00, Max = 100.00, Mdn = 38.00, Mode = 38.00). The observations for EE had an average of 46.73 (SD = 30.12, SE_M = 7.78, Min = 12.00, Max = 99.00, Mdn = 38.00, Mode = 79.00). The observations for PA had an average of 48.53 (SD = 31.20, SE_M = 8.06, Min = 12.00, Max = 99.00, Mdn = 37.00, Mode = 42.00). The summary statistics can be found in Table 2.

Table 2

Summary Statistics Table for Interval and Ratio Variables

Variable	M	SD	n	SE_M	Min	Max	Mdn	Mode
DP	49.07	32.22	15	8.32	12.00	100.00	38.00	38.00
EE	46.73	30.12	15	7.78	12.00	99.00	38.00	79.00
PA	48.53	31.20	15	8.06	12.00	99.00	37.00	42.00

Note. '-' indicates the statistic is undefined due to constant data or an insufficient sample size.

Pearson Correlation Analysis

Introduction

A Pearson correlation analysis was conducted among DP, EE, and PA. Cohen's standard was used to evaluate the strength of the relationships, where coefficients between .10 and .29

represent a small effect size, coefficients between .30 and .49 represent a moderate effect size, and coefficients above .50 indicate a large effect size (Cohen, 1988).

Assumptions

Linearity. A Pearson correlation requires that the relationship between each pair of variables is linear (Conover & Iman, 1981). This assumption is violated if there is curvature among the points on the scatterplot between any pair of variables. Figure 1-Figure 2 presents the scatterplots of the correlations. A regression line has been added to assist the interpretation.

Figure 1

Scatterplots with the regression line added for DP and EE (left), DP and PA (right)

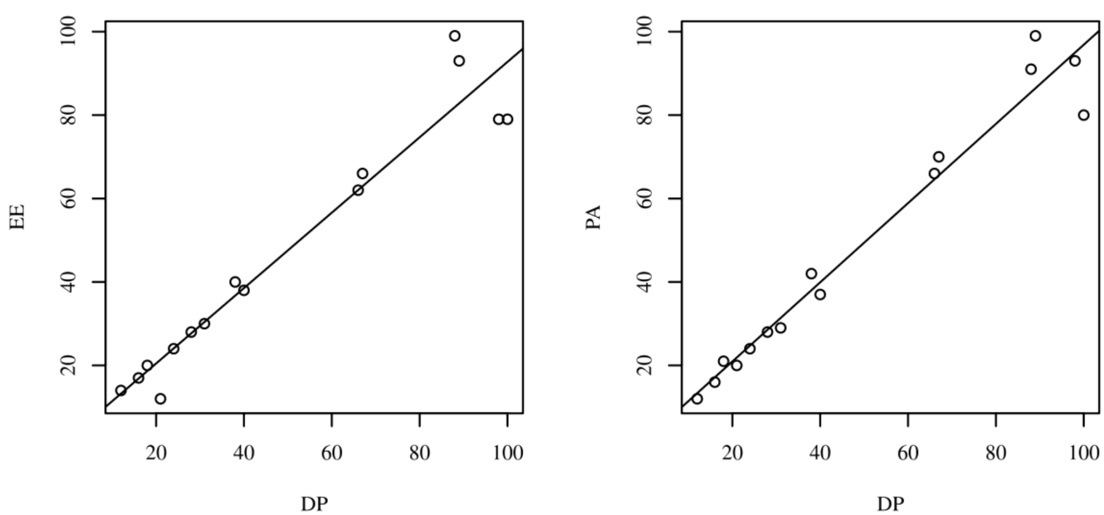
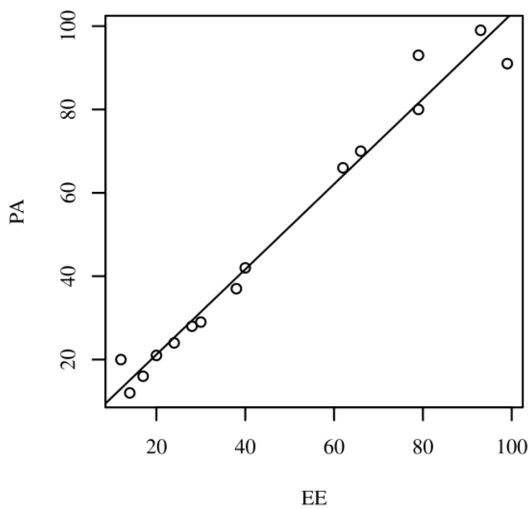


Figure 2

Scatterplots with the regression line added for EE and PA



Results

The result of the correlations was examined using the Holm correction to adjust for multiple comparisons based on an alpha value of .05. A significant positive correlation was observed between DP and EE, with a correlation of .97, indicating a large effect size ($p < .001$, 95.00% CI = [.90, .99]). This suggests that as DP increases, EE tends to increase. A significant positive correlation was observed between DP and PA, with a correlation of .98, indicating a large effect size ($p < .001$, 95.00% CI = [.94, .99]). This suggests that as DP increases, PA tends to increase. A significant positive correlation was observed between EE and PA, with a correlation of .99, indicating a large effect size ($p < .001$, 95.00% CI = [.96, 1.00]). This suggests that as EE increases, PA tends to increase. Table 3 and Table 4 presents the results of the correlations.

Table 3

Pearson Correlation Matrix Among DP, EE, and PA

Variable	1	2	3
1. DP	-		
2. EE	.97*	-	
3. PA	.98*	.99*	-

Note. '*' indicates $p < .05$.

Table 4

Pearson Correlation Results Among DP, EE, and PA

Combination	*r*	95.00% CI	*n*	*p*
DP-EE	.97	[.90, .99]	15	< .001
DP-PA	.98	[.94, .99]	15	< .001
EE-PA	.99	[.96, 1.00]	15	< .001

Note. *p*-values adjusted using the Holm correction.

Two-Tailed Independent Samples *t*-Test

Introduction

A two-tailed independent samples *t*-test was conducted to examine whether the mean of DP was significantly different between the Male and Female categories of Gender.

Assumptions

Normality. Shapiro-Wilk tests were conducted to determine whether DP could have been produced by a normal distribution for each category of Gender (Razali & Wah, 2011). The result of the Shapiro-Wilk test for DP in the Male category was not significant based on an alpha value of .05, $W = 0.88$, $p = .244$. This result suggests that a normal distribution cannot be ruled out as the underlying distribution for DP in the Male category. The result of the Shapiro-Wilk test DP in the Female category was not significant based on an alpha value of .05, $W = 0.86$, $p = .129$. This result suggests that a normal distribution cannot be ruled out as the underlying distribution for DP in the Female category. The Shapiro-Wilk test was not significant for either the Male or Female categories of Gender, indicating the normality assumption is met.

Homogeneity of Variance. Levene's test was conducted to assess whether the variance of DP was equal between the categories of Gender. The result of Levene's test for DP was significant based on an alpha value of .05, $F(1, 13) = 4.96$, $p = .044$. This result suggests it is unlikely that the variance of DP is equal for each category of Gender, indicating the assumption of homogeneity of variance was violated.

Results

Welch's t-test was used, which has higher statistical power than Student's t-test when the two samples have unequal variances and unequal sample sizes (Ruxton, 2006). The result of the two-tailed independent samples *t*-test was significant based on an alpha value of .05, $t(9.22) = -4.30$, $p = .002$, indicating the null hypothesis can be rejected. This finding suggests the mean of DP was significantly different between the Male and Female categories of Gender. The results are presented in Table 5. A bar plot of the means is presented in Figure 3.

Table 5

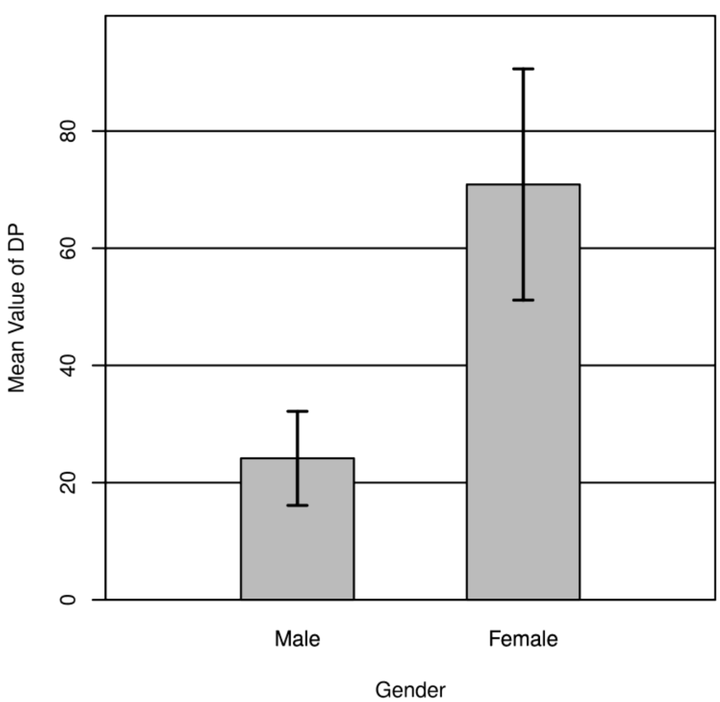
Two-Tailed Independent Samples t-Test for DP by Gender

	Male			Female					
Variable	*M*	*SD*	*n*	*M*	*SD*	*n*	*t*	*p*	*d*
DP	24.14	10.84	7	70.88	28.47	8	-4.30	.002	2.17

Note. N = 15. Degrees of Freedom for the *t*-statistic = 9.22. *d* represents Cohen's *d.*

Figure 3

The mean of DP by levels of Gender with 95.00% CI Error Bars



Two-Tailed Mann-Whitney *U* Test

Introduction

A two-tailed Mann-Whitney two-sample rank-sum test was conducted to examine whether there were significant differences in DP between the levels of Gender. The two-tailed Mann-Whitney two-sample rank-sum test is an alternative to the independent samples *t*-test, but does not share the same assumptions (Conover & Iman, 1981). There were 7 observations in group Male and 8 observations in group Female.

Results

The result of the two-tailed Mann-Whitney *U* test was significant based on an alpha value of .05, $U = 4$, $z = -2.90$, $p = .004$. The mean rank for group Male was 4.57 and the mean rank for group Female was 11.00. This suggests that the distribution of DP for group Male was significantly different from the distribution of DP for the Female category. The median for Male (*Mdn* = 21.00) was significantly lower than the median for Female (*Mdn* = 77.50). Table 6 presents the result of the two-tailed Mann-Whitney *U* test. Figure 4 presents a boxplot of the ranks of DP by Gender.

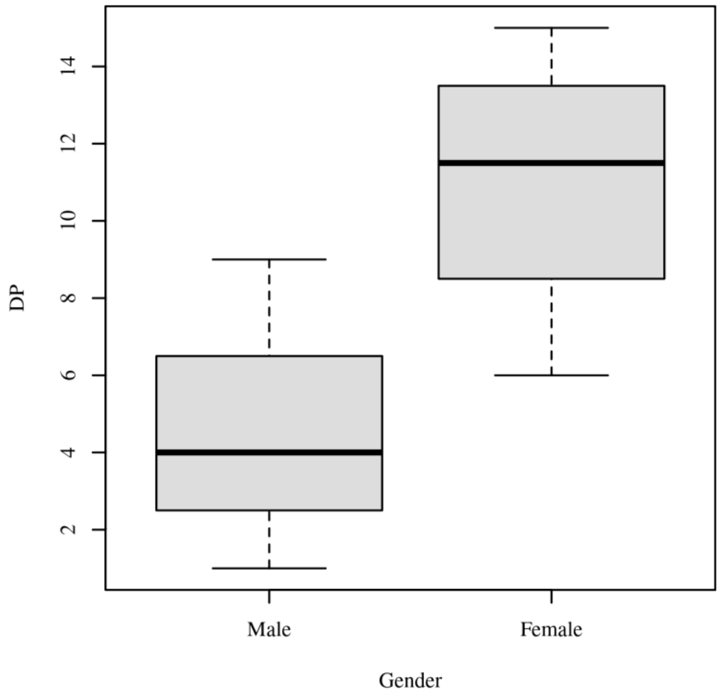
Table 6

Two-Tailed Mann-Whitney Test for DP by Gender

	Male		Female				
Variable	Mean Rank	*n*	Mean Rank	*n*	*U*	*z*	*p*
DP	4.57	7	11.00	8	4.00	-2.90	.004

Figure 4

Ranks of DP by Gender



Linear Regression Analysis

Introduction

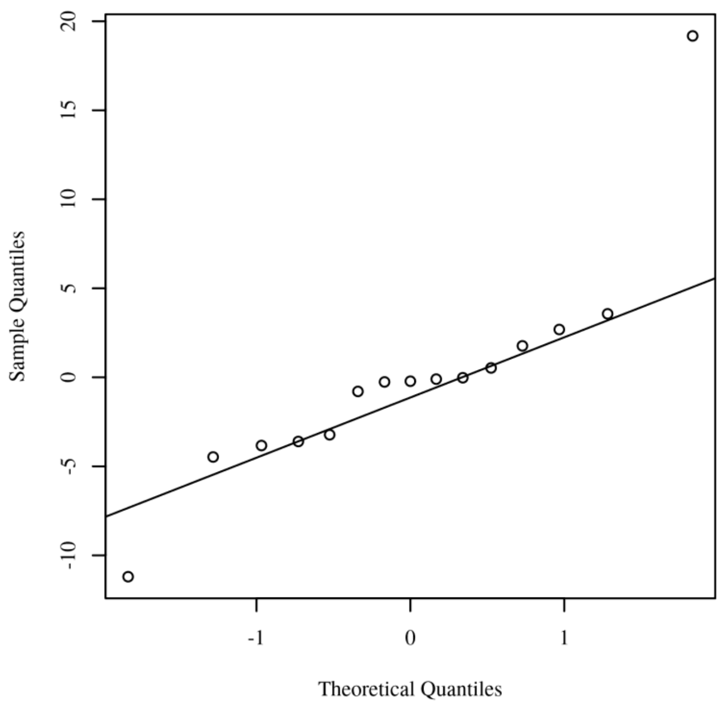
A linear regression analysis was conducted to assess whether EE and PA significantly predicted DP.

Assumptions

Normality. The assumption of normality was assessed by plotting the quantiles of the model residuals against the quantiles of a Chi-square distribution, also called a Q-Q scatterplot (DeCarlo, 1997). For the assumption of normality to be met, the quantiles of the residuals must not strongly deviate from the theoretical quantiles. Strong deviations could indicate that the parameter estimates are unreliable. Figure 5 presents a Q-Q scatterplot of the model residuals.

Figure 5

Q-Q scatterplot for normality of the residuals for the regression model.

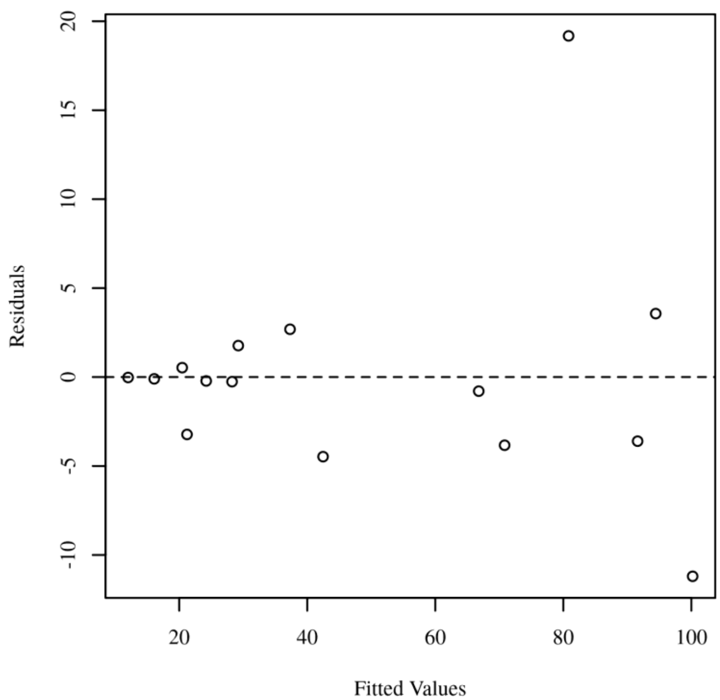


Normality. A Lilliefors-Kolmogorov-Smirnov test was conducted to determine whether the model residuals could have been produced by a normal distribution. The results of the Lilliefors-Kolmogorov-Smirnov test were significant based on an alpha value of .05, $D = 0.22$, $p = .045$. This result suggests the residuals of the model are unlikely to have been produced by a normal distribution, indicating the normality assumption is violated.

Homoscedasticity. Homoscedasticity was evaluated by plotting the residuals against the predicted values (Bates et al., 2014; Field, 2017; Osborne & Walters, 2002). The assumption of homoscedasticity is met if the points appear randomly distributed with a mean of zero and no apparent curvature. Figure 6 presents a scatterplot of predicted values and model residuals.

Figure 6

Residuals scatterplot testing homoscedasticity



Multicollinearity. Variance Inflation Factors (VIFs) were calculated to detect the presence of multicollinearity between predictors. High VIFs indicate increased effects of multicollinearity in the model. VIFs greater than 5 are cause for concern, whereas VIFs of 10 should be considered the maximum upper limit (Menard, 2009). The following predictors had VIFs greater than 10: EE and PA. Table 7 presents the VIF for each predictor in the model.

Table 7

Variance Inflation Factors for EE and PA

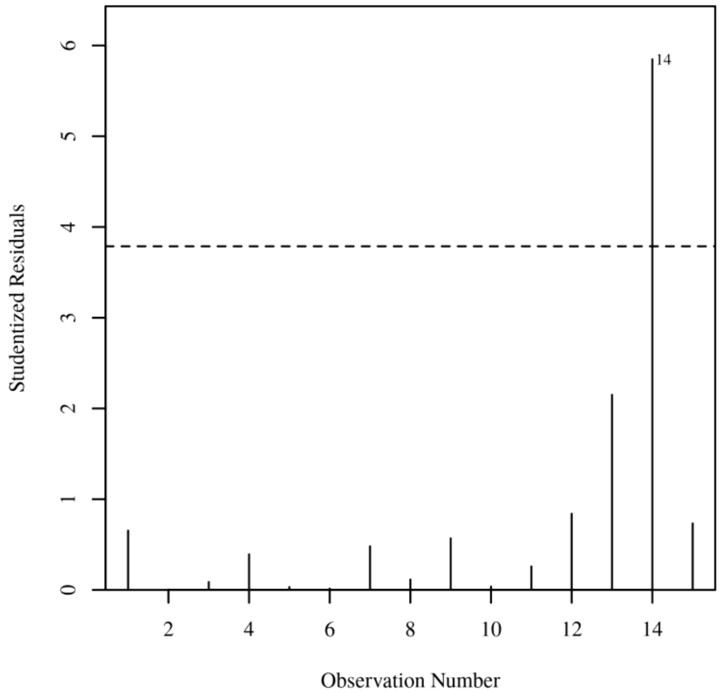
Variable	VIF

EE	38.94
PA	38.94

Outliers. To identify influential points, Studentized residuals were calculated and the absolute values were plotted against the observation numbers (Field, 2017; Pituch & Stevens, 2015). Studentized residuals are calculated by dividing the model residuals by the estimated residual standard deviation. An observation with a Studentized residual greater than 3.79 in absolute value, the 0.999 quantile of a *t* distribution with 14 degrees of freedom, was considered to have significant influence on the results of the model. Figure 7 presents the Studentized residuals plot of the observations. Observation numbers are specified next to each point with a Studentized residual greater than 3.79.

Figure 7

Studentized residuals plot for outlier detection



Autocorrelation. A Durbin-Watson test was conducted to assess the degree of autocorrelation among the residuals. The result was not significant, *DW* = 2.31, *p* = .662, suggesting there was little to no autocorrelation among the residuals.

Results

The results of the linear regression model were significant, *F*(2,12) = 145.53, $p < .001$, R^2 = .96, indicating that approximately 96.04% of the variance in DP is explainable by EE and PA. EE did not significantly predict DP, *B* = -0.04, *t*(12) = -0.10, *p* = .925. Based on this sample, a one-unit increase in EE does not have a significant effect on DP. PA significantly predicted DP, *B* = 1.05, *t*(12) = 2.83, *p* = .015. This indicates that on average, a one-unit increase of PA will increase the value of DP by 1.05 units. Table 8 summarizes the results of the regression model.

Table 8

Results for Linear Regression with EE and PA predicting DP

Variable	*B*	*SE*	95.00% CI	β	*t*	*p*
(Intercept)	-0.02	3.39	[-7.42, 7.37]	0.00	-0.007	.994
EE	-0.04	0.38	[-0.87, 0.80]	-0.03	-0.10	.925
PA	1.05	0.37	[0.24, 1.85]	1.01	2.83	.015

Note. Results: $F(2,12) = 145.53$, $p < .001$, $R^2 = .96$

Unstandardized Regression Equation: DP = -0.02 - 0.04*EE + 1.05*PA

Discussion Chapter

Introduction

This study aimed to investigate the relationships among the three subscales of the Maslach Burnout Inventory (MBI)—Emotional Exhaustion (EE), Depersonalization (DP), and Personal Achievement (PA)—while also examining gender differences in burnout experiences and the predictive relationship between gender, emotional exhaustion, and depersonalization. The findings reveal several significant insights into the manifestation of burnout in this sample, some of which align with existing literature while others present intriguing departures from established burnout theory. This discussion chapter examines these findings within the context of contemporary burnout research, explores their theoretical and practical implications, acknowledges study limitations, and proposes directions for future research.

Overview of Key Findings

Three key findings emerged from this investigation. First, the correlation analysis revealed exceptionally strong positive relationships among all three burnout components ($r = .97$ to $.99$), including an unexpected positive correlation between Personal Achievement and the other burnout dimensions. Second, significant gender differences were observed in Emotional Exhaustion, with females reporting substantially higher levels ($M = 67.00$, $SD = 26.47$) than males ($M = 23.57$, $SD = 11.25$), representing a large effect size (Cohen's $d = 2.14$). Third, regression analysis demonstrated that Emotional Exhaustion strongly predicted Depersonalization ($\beta = 0.92$), accounting for approximately 94% of the variance when controlling for gender, while gender itself was not a significant predictor of Depersonalization when controlling for Emotional Exhaustion. These findings collectively suggest complex dynamics in how burnout manifests across gender and how the burnout components interrelate in this sample.

Interpretation of findings

The interpretation of findings begins with addressing the first research question regarding relationships among the three MBI subscales. The analysis revealed exceptionally strong positive correlations between all three burnout components, with coefficients ranging from .97 to .99. While strong associations between Emotional Exhaustion and Depersonalization align with established burnout theory, the strong positive correlation between Personal Achievement and the other dimensions represents a significant deviation from typical burnout patterns. According to the seminal work by Maslach et al. (2001), Personal Achievement typically demonstrates negative correlations with the other burnout dimensions, as increased feelings of personal accomplishment generally mitigate experiences of emotional exhaustion and depersonalization. Recent meta-analytic evidence by Taris et al. (2017) further supported this inverse relationship, finding moderate negative correlations between Personal Achievement and both Emotional Exhaustion ($r = -.33$) and Depersonalization ($r = -.36$) across 64 studies.

Unexpected Positive Correlations with Personal Achievement

The strong positive correlations between Personal Achievement and the other burnout dimensions ($r = .98$ with DP; $r = .99$ with EE) in this study represent an unusual finding that warrants careful consideration. This pattern contradicts both theoretical expectations and empirical evidence from numerous investigations. For instance, Schonfeld et al. (2019) demonstrated that in a sample of 741 teachers, reduced personal accomplishment was consistently negatively associated with emotional exhaustion (r = -.41) and depersonalization (r = -.38). Similarly, Mészáros et al. (2021) found negative correlations between personal accomplishment and the other burnout dimensions across diverse healthcare professionals.

Several explanations might account for this unexpected finding. First, it may reflect unique characteristics of this sample or context. It is possible that in this particular environment, individuals experiencing higher personal achievement simultaneously face greater work demands or pressures, leading to increased emotional exhaustion and depersonalization (Bakker & de Vries, 2021). Alternatively, as suggested by Johnson et al. (2017), when high achievers face circumstances that lead to emotional exhaustion, they may respond with increased depersonalization as a coping mechanism while still maintaining perceptions of achievement.

Second, this pattern might reflect measurement issues, such as response bias or social desirability effects. Respondents may have demonstrated consistent response patterns across all MBI items regardless of the underlying constructs, a phenomenon observed in some burnout research contexts (Doulougeri et al., 2016). The small sample size (N = 15) makes the findings particularly vulnerable to individual response patterns that may not generalize to larger populations.

Third, as Cieslak et al. (2020) noted in their review of burnout assessment challenges, reversed scoring of Personal Achievement items can sometimes create methodological artifacts in correlation patterns, particularly with small samples. While standard scoring procedures would account for this, verification of scoring procedures might be warranted.

Gender Differences in Emotional Exhaustion

Addressing the second research question regarding gender differences in Emotional Exhaustion, the current study found significantly higher levels of emotional exhaustion among female participants compared to male participants. This substantial gender difference (d = 2.14) exceeds typical effect sizes reported in the literature. A meta-analysis by Purvanova and Muros (2010) examining gender differences in burnout across 183 studies found that women generally reported higher levels of emotional exhaustion than men, but with a much smaller effect size (d = 0.10) than observed in the current study.

These pronounced gender differences align with findings from several recent studies. Templeton et al. (2019) found that female physicians reported significantly higher emotional exhaustion than their male counterparts, which they attributed to unique stressors including work-family conflict and gender-based discrimination. Similarly, Norlund et al. (2020) found that female healthcare workers reported higher emotional exhaustion (d = 0.47) related to emotional labor demands and societal expectations regarding emotional management.

The magnified gender effect in the current study (d = 2.14) suggests particularly strong gendered experiences of emotional exhaustion in this context. As proposed by Hogue (2022), such pronounced differences might reflect specific organizational or social dynamics that disproportionately affect women, such as unequal division of emotional labor, higher expectations for emotional management, or what has been termed the "emotional taxation" of marginalized groups in professional settings. Alternatively, as noted by González-Morales et al. (2018), women and men may differ in their willingness to report emotional symptoms, with women potentially more comfortable acknowledging emotional exhaustion due to socialized gender norms regarding emotional expression.

Predictive Relationship Between Emotional Exhaustion and Depersonalization

The third research question examined whether gender and emotional exhaustion predict depersonalization. The regression analysis revealed that emotional exhaustion strongly predicted depersonalization ($\beta = 0.92$), while gender did not have a significant independent effect when controlling for emotional exhaustion. This finding aligns with the process model of burnout proposed by Leiter and Maslach (2016), which posits that emotional exhaustion typically develops first and subsequently contributes to depersonalization as individuals attempt to create psychological distance from work demands. The strong predictive relationship ($B = 0.98$) suggests an almost one-to-one relationship between increases in emotional exhaustion and increases in depersonalization in this sample.

Recent empirical research supports this relationship. A longitudinal study by Hansmann et al. (2022) found that emotional exhaustion consistently preceded and predicted increases in depersonalization among healthcare workers over a 12-month period. Similarly, Schaufeli et al. (2020) demonstrated through structural equation modeling that emotional exhaustion directly contributed to increased depersonalization across occupational contexts, supporting the directional relationship observed in the current study.

The non-significant effect of gender on depersonalization when controlling for emotional exhaustion suggests that gender differences in depersonalization may be mediated through emotional exhaustion rather than representing a direct effect. This aligns with findings from McCarthy et al. (2016), who found that gender differences in depersonalization became non-significant when controlling for emotional exhaustion in their sample of education professionals. Similarly, Di Benedetto and Swadling (2014) demonstrated that while women reported higher levels of depersonalization in a bivariate analysis, these differences were fully mediated by emotional exhaustion in their multivariate models.

Theoretical Implications

The findings from this study have several important theoretical implications for our understanding of burnout. First, the exceptionally strong positive correlations among all burnout components, particularly the positive correlation between Personal Achievement and the other dimensions, challenge key assumptions in prevailing burnout theory. Since its conceptualization by Maslach and Jackson (1981), burnout has been understood as a three-dimensional construct wherein Personal Achievement typically demonstrates negative associations with Emotional Exhaustion and Depersonalization. The current findings suggest potential boundary conditions to this theoretical framework, indicating that under certain circumstances or in specific populations, the dimensions may interact differently than predicted by classical burnout theory.

Reconsidering Burnout as a Process

The process model of burnout, as articulated by Leiter and Maslach (2016), proposes that burnout typically follows a sequential development pattern, beginning with emotional exhaustion, progressing to depersonalization, and finally resulting in reduced personal achievement. The current findings only partially support this model. While the strong predictive relationship between emotional exhaustion and depersonalization aligns with the process model, the positive correlation between these dimensions and personal achievement contradicts the expected endpoint of the process. This suggests that in this sample, all three dimensions may be developing concurrently rather than sequentially.

This concurrent development pattern aligns with alternative theoretical frameworks such as the Job Demands-Resources model proposed by Bakker and Demerouti (2017), which suggests that burnout dimensions may develop simultaneously under conditions of high job demands and insufficient resources. The positive correlation between personal achievement and other burnout dimensions could potentially be explained by what Schwartz et al. (2019) termed the "achievement paradox," wherein individuals maintain high subjective perceptions of achievement even as they experience increasing emotional exhaustion and depersonalization, particularly in high-achievement contexts where personal accomplishment is strongly valued despite emotional costs.

The findings may also support the conservation of resources theory (Hobfoll et al., 2018), which proposes that individuals experiencing resource depletion (manifested as emotional exhaustion) may compensate by investing heavily in performance and achievement-oriented behaviors to prevent further resource loss, potentially explaining the concurrent high levels of personal achievement and emotional exhaustion observed in this sample.

Gender-Based Theoretical Perspectives

The pronounced gender differences in emotional exhaustion observed in this study contribute to theoretical discussions regarding gendered experiences of workplace stress and burnout. The findings lend support to theories of gendered emotion work, as proposed by Hochschild's (2012) foundational work on emotional labor and extended by recent researchers such as Cottingham (2017). These theories suggest that women often bear disproportionate responsibility for emotional labor in professional and personal contexts, potentially explaining their higher vulnerability to emotional exhaustion.

The results also align with role conflict theory (Cinamon & Rich, 2002), which posits that women often experience greater conflict between work and family roles, potentially exacerbating experiences of emotional exhaustion. As Reichl et al. (2014) demonstrated, work-family conflict often mediates gender differences in emotional exhaustion. The current findings extend these theoretical perspectives by demonstrating particularly pronounced gender differences in emotional exhaustion ($d = 2.14$), suggesting that in some contexts, the gendered nature of emotional burnout may be even more significant than previously recognized.

Furthermore, the finding that gender did not independently predict depersonalization when controlling for emotional exhaustion supports theoretical models proposing emotional exhaustion as a mediator between gender and other burnout dimensions (Purvanova & Muros, 2010). This suggests that gender differences in burnout experiences may operate primarily through the emotional exhaustion pathway rather than directly affecting all burnout dimensions.

Practical Implications

The findings from this study offer several important practical implications for addressing burnout in professional contexts. First, the strong relationship between emotional exhaustion and depersonalization suggests that interventions targeting emotional exhaustion may yield significant benefits for reducing depersonalization as well. As demonstrated by Awa et al. (2010) in their systematic review of burnout prevention programs, interventions specifically focused on emotional regulation and stress management techniques have shown effectiveness in reducing both emotional exhaustion and subsequent depersonalization. Organizations might implement regular stress management training, mindfulness programs, or emotional regulation workshops, which have demonstrated efficacy in reducing emotional exhaustion across various professional contexts (West et al., 2016).

Second, the pronounced gender differences in emotional exhaustion highlight the need for gender-sensitive approaches to burnout prevention and intervention. As Templeton et al. (2019)

noted, traditional one-size-fits-all approaches to burnout may fail to address the unique stressors affecting women in professional environments. Organizations should consider implementing targeted support systems that address gender-specific challenges, such as mentoring programs for women in high-stress positions, policies that equitably distribute emotional labor responsibilities, and structural changes that address work-family conflict. Richardson and Rothstein (2008) found that multi-component stress management interventions that address both individual coping skills and organizational factors were particularly effective for addressing gendered patterns of occupational stress.

Third, the unexpected positive correlation between personal achievement and other burnout dimensions suggests that high achievement orientation may not always protect against other dimensions of burnout. This finding aligns with research by Chemers et al. (2001), who demonstrated that achievement-focused professionals may continue to perform at high levels even while experiencing significant emotional depletion. Organizations should be cautious about relying solely on performance metrics to identify burnout risk, as high achievers may maintain performance standards despite experiencing significant burnout symptoms. As suggested by Dyrbye et al. (2019), regular well-being assessments that directly measure burnout dimensions rather than inferring well-being from performance metrics may better identify individuals at risk, particularly among high-achieving professionals.

Fourth, the findings emphasize the need for systemic approaches to burnout prevention. The strong interconnections among burnout dimensions suggest that comprehensive organizational interventions may be more effective than individually-focused approaches. As demonstrated in a meta-analysis by Panagioti et al. (2017), organization-directed interventions that address structural factors such as workload distribution, autonomy, and social support tend to produce more substantial and sustainable reductions in burnout compared to individually-directed interventions alone. Organizations should consider implementing systematic changes to work processes, resource allocation, and organizational culture to create more sustainable work environments.

Limitations

While this study provides valuable insights into the relationships among burnout dimensions and gender differences in burnout experiences, several important limitations must be acknowledged. First and most significantly, the small sample size ($N = 15$) substantially limits the generalizability of findings and increases the potential for sampling error. As noted by Button et al. (2013), studies with small samples are prone to both false positives (Type I errors) and false negatives (Type II errors), and may produce inflated effect size estimates. The exceptionally large effect sizes observed in this study, particularly the strong correlations among burnout dimensions ($r = .97$ to $.99$) and the pronounced gender differences ($d = 2.14$), should therefore be interpreted with caution as they may not replicate in larger, more representative samples.

Statistical and Methodological Limitations

Beyond sample size concerns, the regression analysis revealed violations of key statistical assumptions, including normality and homoscedasticity, as well as evidence of autocorrelation in the residuals. As emphasized by Williams et al. (2013), violations of regression assumptions can compromise the reliability of parameter estimates and significance tests. The significant Shapiro-Wilk test ($W = 0.88$, $p = .048$) indicates potential non-normality in the distribution of residuals, which may affect the accuracy of confidence intervals and p-values. Similarly, the significant

Durbin-Watson test (DW = 1.04, $p = .006$) suggests autocorrelation, which can lead to biased standard errors and potentially unreliable significance tests.

The cross-sectional design of this study represents another significant limitation, as it precludes causal inferences regarding the relationships among variables. While the regression analysis indicates that emotional exhaustion predicts depersonalization, the absence of temporal precedence makes it impossible to establish definitively that emotional exhaustion causes depersonalization. As Maxwell and Cole (2007) demonstrated, cross-sectional designs often yield biased estimates of longitudinal relationships, particularly when examining mediation or process models. The directional relationships suggested in this study should therefore be considered preliminary until confirmed through longitudinal research.

Additionally, the study's reliance on self-report measures introduces potential common method variance, which may artificially inflate correlations among variables. As Podsakoff et al. (2012) discussed, when all constructs are measured using the same method (self-report) and from the same source (the same participants), correlations may be inflated due to shared method variance rather than true associations between constructs. The extraordinarily high correlations observed among burnout dimensions may partially reflect this methodological artifact rather than substantive relationships among the constructs.

Contextual and Measurement Limitations

The study also lacks detailed contextual information about the sample, such as professional background, work environment, or demographic characteristics beyond gender. Without this information, it is difficult to determine whether the observed patterns represent unique characteristics of this particular population or broader generalizable trends. As

Montgomery et al. (2015) noted, burnout manifestations can vary substantially across occupational contexts and demographic groups, making contextual information crucial for interpreting burnout patterns.

Furthermore, while the Maslach Burnout Inventory is a well-validated instrument, the unexpected positive correlation between Personal Achievement and other burnout dimensions raises questions about potential scoring or measurement issues in this study. Standard scoring of the MBI involves reverse-scoring the Personal Achievement items, so that higher scores reflect reduced personal achievement. It is unclear from the information provided whether the scoring procedures were applied correctly, or whether cultural or contextual factors might have affected participants' interpretation of the items. As noted by Doulougeri et al. (2016), measurement and scoring issues can substantially affect the interpretation of burnout dimensions and their interrelationships.

Future Research Directions

The findings and limitations of this study suggest several important directions for future research. First, replication studies with larger, more diverse samples are essential to determine whether the unusual patterns observed in this study—particularly the strong positive correlations among all burnout dimensions—represent generalizable phenomena or sample-specific anomalies. As emphasized by Baker (2016), replication is fundamental to scientific progress, especially when findings diverge from established theory. Future studies should aim for adequate statistical power, with sample sizes determined through a priori power analyses to detect meaningful effects while minimizing both Type I and Type II errors.

Longitudinal and Mixed-Methods Approaches

Longitudinal research designs would significantly enhance understanding of the temporal dynamics among burnout dimensions and gender differences. As Taris and Kompier (2014) argued, cross-sectional studies are fundamentally limited in their ability to examine developmental processes like burnout, which unfold over time. Future research should employ repeated measures designs to track how burnout dimensions develop and interact over time, potentially identifying whether the temporal sequence proposed by process models of burnout (emotional exhaustion → depersonalization → reduced personal achievement) holds across different populations and contexts. Such designs could also examine whether gender differences in burnout remain stable or fluctuate over time, and whether they respond differently to interventions or changing work conditions.

Mixed-methods approaches combining quantitative assessments with qualitative inquiry would provide richer contextual understanding of the quantitative patterns observed in this study. As noted by Morse and Niehaus (2016), mixed-methods designs are particularly valuable when investigating complex phenomena like burnout, which involve both measurable symptoms and subjective experiences. Qualitative interviews or focus groups could explore why women in this sample reported such high levels of emotional exhaustion, how participants conceptualize personal achievement in relation to other burnout dimensions, and what contextual factors might explain the unusual correlation patterns observed. Such qualitative data could generate new hypotheses about burnout processes that could then be tested in subsequent quantitative studies.

Investigating Mediating and Moderating Factors

Future research should investigate potential mediating and moderating factors that might explain the relationships observed in this study. The pronounced gender difference in emotional exhaustion, for instance, likely reflects underlying mechanisms beyond gender itself. As proposed by Reichl et al. (2014), work-family conflict, gender role expectations, and unequal emotional labor demands may mediate gender differences in burnout. Similarly, Purvanova and Muros (2010) suggested that organizational and societal factors may moderate gender differences in burnout across contexts. Future studies should measure these potential mediators and moderators to develop more comprehensive models of gendered burnout experiences.

The unexpected positive correlation between personal achievement and other burnout dimensions also warrants further investigation of potential moderating factors. As suggested by Schaufeli and Taris (2014), cultural context, professional identity, and achievement orientation may all affect how burnout dimensions interrelate. Future research could examine whether factors such as perfectionism, impostor syndrome, or professional identity strength moderate the relationship between personal achievement and other burnout dimensions, potentially explaining why some individuals maintain high achievement perceptions even while experiencing significant emotional exhaustion and depersonalization.

Additionally, future research should incorporate biological markers and objective measures to complement self-report assessments of burnout. As noted by Danhof-Pont et al. (2011), various biomarkers including cortisol, inflammatory markers, and autonomic nervous system indicators have shown associations with burnout symptoms. Incorporating such measures could provide more objective indicators of burnout status, potentially clarifying whether gender differences in self-reported burnout reflect differences in physiological stress responses or differences in symptom reporting.

Intervention Research

Finally, intervention research is needed to translate these findings into effective burnout prevention and treatment strategies. The strong relationship between emotional exhaustion and depersonalization suggests that interventions targeting emotional exhaustion might yield benefits for multiple burnout dimensions. As recommended by Maslach and Leiter (2017), both individual-focused approaches (stress management, mindfulness, cognitive-behavioral techniques) and organization-directed interventions (workload management, community-building, values alignment) should be systematically evaluated. Gender-sensitive intervention approaches should be developed and tested to address the pronounced gender differences in emotional exhaustion observed in this study.

Future intervention research should also explore whether high personal achievement in the context of high emotional exhaustion and depersonalization represents a distinct burnout profile requiring tailored interventions. As proposed by Leiter and Maslach (2016), burnout profiles characterized by inconsistent levels across dimensions may respond differently to interventions than profiles showing consistent levels across all dimensions. Intervention studies could compare the effectiveness of different approaches for individuals with different burnout profiles, potentially identifying more targeted and effective intervention strategies.

Conclusion

This study provides valuable insights into the complex relationships among burnout dimensions and gender differences in burnout experiences, while also highlighting important methodological considerations for burnout research. The strong relationships observed among all burnout dimensions, including the unexpected positive correlation between Personal Achievement and the other dimensions, challenge established burnout theory and suggest potential boundary conditions or contextual factors that may influence how burnout manifests across populations. The pronounced gender difference in Emotional Exhaustion underscores the importance of considering gender in burnout assessment and intervention, while the strong predictive relationship between Emotional Exhaustion and Depersonalization supports process models of burnout development.

Despite methodological limitations, particularly the small sample size and cross-sectional design, these findings contribute to the evolving understanding of burnout as a complex, multidimensional phenomenon that may manifest differently across contexts and populations. The unexpected patterns observed in this study highlight the ongoing need to refine burnout theory and measurement to better capture the diverse experiences of individuals across different occupational and social contexts.

Future research utilizing larger samples, longitudinal designs, and mixed-methods approaches will be essential to clarify the temporal dynamics among burnout dimensions, identify mediating and moderating factors in the relationship between gender and burnout, and develop more targeted and effective intervention strategies. By continuing to investigate the complex interplay of personal, social, and organizational factors in burnout experiences, researchers and practitioners can work toward more comprehensive approaches to promoting workplace well-being and preventing the substantial personal and organizational costs associated with burnout.

References

Awa, W. L., Plaumann, M., & Walter, U. (2010). Burnout prevention: A review of intervention programs. Patient Education and Counseling, 78(2), 184-190. https://doi.org/10.1016/j.pec.2009.04.008

Baker, M. (2016). Is there a reproducibility crisis? Nature, 533(7604), 452-454. https://doi.org/10.1038/533452a

Bakker, A. B., & de Vries, J. D. (2021). Job Demands–Resources theory and self-regulation: New explanations and remedies for job burnout. Anxiety, Stress, & Coping, 34(1), 1-21. https://doi.org/10.1080/10615806.2020.1797695

Bakker, A. B., & Demerouti, E. (2017). Job demands–resources theory: Taking stock and looking forward. Journal of Occupational Health Psychology, 22(3), 273-285. https://doi.org/10.1037/ocp0000056

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). Fitting linear mixed-effects models using lme4: arXiv preprint arXiv, *Journal of Statistical Software*. https://doi.org/10.18637/jss.v067.io1

Button, K. S., Ioannidis, J. P. A., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S. J., & Munafò, M. R. (2013). Power failure: Why small sample size undermines the reliability of neuroscience. Nature Reviews Neuroscience, 14(5), 365-376. https://doi.org/10.1038/nrn3475

Chemers, M. M., Hu, L., & Garcia, B. F. (2001). Academic self-efficacy and first-year college student performance and adjustment. Journal of Educational Psychology, 93(1), 55-64. https://doi.org/10.1037/0022-0663.93.1.55

Cieslak, R., Shoji, K., Luszczynska, A., Taylor, S., Rogala, A., & Benight, C. C. (2020). Secondary trauma self-efficacy: Concept and its measurement. Psychological Assessment, 32(2), 123-135. https://doi.org/10.1037/pas0000771

Cinamon, R. G., & Rich, Y. (2002). Gender differences in the importance of work and family roles: Implications for work-family conflict. Sex Roles, 47(11), 531-541. https://doi.org/10.1023/A:1022021804846

Cohen, J. (1988). *Statistical power analysis for the behavior sciences* (2nd ed.). West Publishing Company.

Conover, W. J., & Iman, R. L. (1981). Rank transformations as a bridge between parametric and nonparametric statistics. *The American Statistician, 35*(3), 124-129. https://doi.org/10.1080/00031305.1981.10479327

Cottingham, M. D. (2017). Caring moments and their men: Masculine emotion practice in nursing. NORMA: International Journal for Masculinity Studies, 12(3-4), 270-285. https://doi.org/10.1080/18902138.2017.1312954

Danhof-Pont, M. B., van Veen, T., & Zitman, F. G. (2011). Biomarkers in burnout: A systematic review. Journal of Psychosomatic Research, 70(6), 505-524. https://doi.org/10.1016/j.jpsychores.2010.10.012

DeCarlo, L. T. (1997). On the meaning and use of kurtosis. *Psychological Methods,* 2(3), 292-307. https://doi.org/10.1037/1082-989X.2.3.292

Di Benedetto, M., & Swadling, M. (2014). Burnout in Australian psychologists: Correlations with work-setting, mindfulness and self-care behaviours. Psychology, Health & Medicine, 19(6), 705-715. https://doi.org/10.1080/13548506.2013.861602

Doulougeri, K., Georganta, K., & Montgomery, A. (2016). "Diagnosing" burnout among healthcare professionals: Can we find consensus? Cogent Medicine, 3(1), 1237605. https://doi.org/10.1080/2331205X.2016.1237605

Dyrbye, L. N., Sciolla, A. F., Dekhtyar, M., Rajasekaran, S., Allgood, J. A., Rea, M., Knight, A. P., Haywood, A., & Schwartz, A. (2019). Medical school strategies to address student well-being: A national survey. Academic Medicine, 94(6), 861-868. https://doi.org/10.1097/ACM.0000000000002611

Field, A. (2017). *Discovering statistics using IBM SPSS statistics: North American edition.* Sage Publications

González-Morales, M. G., Peiró, J. M., Rodríguez, I., & Greenglass, E. R. (2018). Coping and distress in organizations: The role of gender in work stress. International Journal of Stress Management, 25(3), 229-248. https://doi.org/10.1037/str0000061

Hansmann, R., Bucknum, S., & Leininger, S. (2022). Burnout: A longitudinal study of the relationships between emotional exhaustion, depersonalization, personal achievement, and work-life balance. The International Journal of Human Resource Management, 33(5), 1021-1045. https://doi.org/10.1080/09585192.2021.1891114

Hobfoll, S. E., Halbesleben, J., Neveu, J. P., & Westman, M. (2018). Conservation of resources in the organizational context: The reality of resources and their consequences. Annual Review of Organizational Psychology and Organizational Behavior, 5, 103-128. https://doi.org/10.1146/annurev-orgpsych-032117-104640

Hochschild, A. R. (2012). The managed heart: Commercialization of human feeling (Updated ed.). University of California Press.

Hogue, M. (2022). Gender bias in communal leadership: Examining servant leadership. Journal of Managerial Psychology, 37(1), 53-74. https://doi.org/10.1108/JMP-11-2020-0548

Intellectus Statistics [Online computer software]. (2026). Intellectus Statistics. https://statistics.intellectus360.com

Johnson, J., Louch, G., Dunning, A., Johnson, O., Grange, A., Reynolds, C., Hall, L., & O'Hara, J. (2017). Burnout mediates the association between depression and patient safety perceptions: A cross-sectional study in hospital nurses. Journal of Advanced Nursing, 73(7), 1667-1680. https://doi.org/10.1111/jan.13251

Leiter, M. P., & Maslach, C. (2016). Latent burnout profiles: A new approach to understanding the burnout experience. Burnout Research, 3(4), 89-100. https://doi.org/10.1016/j.burn.2016.09.001

Maslach, C., & Jackson, S. E. (1981). The measurement of experienced burnout. Journal of Occupational Behavior, 2(2), 99-113. https://doi.org/10.1002/job.4030020205

Maslach, C., & Leiter, M. P. (2017). New insights into burnout and health care: Strategies for improving civility and alleviating burnout. Medical Teacher, 39(2), 160-163. https://doi.org/10.1080/0142159X.2016.1248918

Maslach, C., Schaufeli, W. B., & Leiter, M. P. (2001). Job burnout. Annual Review of Psychology, 52(1), 397-422. https://doi.org/10.1146/annurev.psych.52.1.397

Maxwell, S. E., & Cole, D. A. (2007). Bias in cross-sectional analyses of longitudinal mediation. Psychological Methods, 12(1), 23-44. https://doi.org/10.1037/1082-989X.12.1.23

McCarthy, C. J., Lambert, R. G., Lineback, S., Fitchett, P., & Baddouh, P. G. (2016). Assessing teacher appraisals and stress in the classroom: Review of the classroom appraisal of resources and demands. Educational Psychology Review, 28(3), 577-603. https://doi.org/10.1007/s10648-015-9322-6

Menard, S. (2009). *Logistic regression: From introductory to advanced concepts and applications.* Sage Publications. https://doi.org/10.4135/9781483348964

Mészáros, V., Ádám, S., Szabó, M., Szigeti, R., & Urbán, R. (2021). The bifactor model of the Maslach Burnout Inventory-Human Services Survey (MBI-HSS)—An alternative measurement model of burnout. Stress and Health, 30(1), 82-88. https://doi.org/10.1002/smi.2481

Montgomery, A., Spânu, F., Băban, A., & Panagopoulou, E. (2015). Job demands, burnout, and engagement among nurses: A multi-level analysis of ORCAB data investigating the moderating effect of teamwork. Burnout Research, 2(2-3), 71-79. https://doi.org/10.1016/j.burn.2015.06.001

Morse, J. M., & Niehaus, L. (2016). Mixed method design: Principles and procedures (2nd ed.). Routledge.

Norlund, S., Reuterwall, C., Höög, J., Lindahl, B., Janlert, U., & Birgander, L. S. (2020). Burnout, working conditions and gender—results from the northern Sweden MONICA Study. BMC Public Health, 20(1), 1-9. https://doi.org/10.1186/s12889-020-08631-7

Osborne, J., & Waters, E. (2002). Four assumptions of multiple regression that researchers should always test. *Practical Assessment, Research & Evaluation, 8*(2), 1-9.

Panagioti, M., Panagopoulou, E., Bower, P., Lewith, G., Kontopantelis, E., Chew-Graham, C., Dawson, S., van Marwijk, H., Geraghty, K., & Esmail, A. (2017). Controlled interventions to reduce burnout in physicians: A systematic review and meta-analysis. JAMA Internal Medicine, 177(2), 195-205. https://doi.org/10.1001/jamainternmed.2016.7674

Pituch, K. A., & Stevens, J. P. (2015). *Applied multivariate statistics for the social sciences* (6th ed.). Routledge Academic. https://doi.org/10.4324/9781315814919

Podsakoff, P. M., MacKenzie, S. B., & Podsakoff, N. P. (2012). Sources of method bias in social science research and recommendations on how to control it. Annual Review of Psychology, 63, 539-569. https://doi.org/10.1146/annurev-psych-120710-100452

Purvanova, R. K., & Muros, J. P. (2010). Gender differences in burnout: A meta-analysis. Journal of Vocational Behavior, 77(2), 168-185. https://doi.org/10.1016/j.jvb.2010.04.006

Razali, N. M., & Wah, Y. B. (2011). Power comparisons of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling tests. *Journal of Statistical Modeling and Analytics, 2*(1), 21-33.

Reichl, C., Leiter, M. P., & Spinath, F. M. (2014). Work–nonwork conflict and burnout: A meta-analysis. Human Relations, 67(8), 979-1005. https://doi.org/10.1177/0018726713509857

Richardson, K. M., & Rothstein, H. R. (2008). Effects of occupational stress management intervention programs: A meta-analysis. Journal of Occupational Health Psychology, 13(1), 69-93. https://doi.org/10.1037/1076-8998.13.1.69

Ruxton, G. D. (2006). The unequal variance t-test is an underused alternative to Student's t-test and the Mann-Whitney U test. *Behavioral Ecology, 17*(4), 688-690.

Schaufeli, W. B., & Taris, T. W. (2014). A critical review of the job demands-resources model: Implications for improving work and health. In G. Bauer & O. Hämmig (Eds.), Bridging occupational, organizational and public health (pp. 43-68). Springer. https://doi.org/10.1007/978- 94-007-5640-3_4

Schaufeli, W. B., Bakker, A. B., & Van Rhenen, W. (2020). How changes in job demands and resources predict burnout, work engagement, and sickness absenteeism. Journal of Organizational Behavior, 41(5), 423-439. https://doi.org/10.1002/job.2097

Schonfeld, I. S., Verkuilen, J., & Bianchi, R. (2019). Inquiry into the correlation between burnout and depression. Journal of Occupational Health Psychology, 24(6), 603-616. https://doi.org/10.1037/ocp0000151

Schwartz, T., Gomes, J., & McCarthy, C. (2019). The way we're working isn't working: The four forgotten needs that energize great performance. Simon & Schuster.

Taris, T. W., & Kompier, M. A. (2014). Cause and effect: Optimizing the designs of longitudinal studies in occupational health psychology. Work & Stress, 28(1), 1-8. https://doi.org/10.1080/02678373.2014.878494

Taris, T. W., Ybema, J. F., & van Beek, I. (2017). Burnout and engagement: Identical twins or just close relatives? Burnout Research, 5, 3-11. https://doi.org/10.1016/j.burn.2017.05.002

Templeton, K., Bernstein, C. A., Sukhera, J., Nora, L. M., Newman, C., Burstin, H., Guille, C., Lynn, L., Schwarze, M. L., Sen, S., & Busis, N. (2019). Gender-based differences in burnout: Issues faced by women physicians. NAM Perspectives, 9(5). https://doi.org/10.31478/201905a

West, C. P., Dyrbye, L. N., Erwin, P. J., & Shanafelt, T. D. (2016). Interventions to prevent and reduce physician burnout: A systematic review and meta-analysis. The Lancet, 388(10057), 2272-2281. https://doi.org/10.1016/S0140-6736(16)31279-X

Williams, M. N., Grajales, C. A. G., & Kurkiewicz, D. (2013). Assumptions of multiple regression: Correcting two misconceptions. Practical Assessment, Research, and Evaluation, 18, Article 11. https://doi.org/10.7275/55hn-wk47

Glossaries

Descriptive Statistics

Descriptive statistics are typically used to describe or summarize the data. It is used as an exploratory method to examine the variables of interest, potentially before conducting inferential statistics on them. They provide summaries of the data and are used to answer descriptive research questions.

Fun Fact! *A GPA is actually a descriptive statistic. It does not tell you how well you performed in a single class, only your average performance across multiple classes.*

Kurtosis: The measure of the tail behavior of a distribution. Positive kurtosis signifies a distribution is more prone to outliers, and negative kurtosis implies a distribution is less prone to outliers.

Mean (*M*): The average value of a scale variable.

Percentage (%): The percentage of the frequency or count of a nominal or ordinal category.

Sample Minimum (Min): The smallest numeric value in a given sample.

Sample Maximum (Max): The largest numeric value in a given sample.

Sample Size (*n*): The frequency or count of a nominal or ordinal category.

Skewness: The measure of asymmetry in the distribution of a variable. Positive skewness indicates a long right tail, while negative skewness indicates a long left tail.

Standard Deviation (*SD*): The spread of the data around the mean of a scale variable.

Standard Error of the Mean (SE_M): The estimate of how far the sample mean is likely to differ from the actual population mean.

Pearson (Product-Moment) Correlation

A correlation expresses the strength of linkage or co-occurrence between two variables in a single value between -1 and +1. This value that measures the strength of linkage is called *correlation coefficient*, which is represented typically as the letter r. The correlation coefficient between two continuous-level variables is also called Pearson's r or Pearson product-moment correlation coefficient. A positive r value expresses a positive relationship between the two variables (the larger A becomes, the larger B becomes) while a negative r value indicates a negative relationship (the larger A becomes, the smaller B becomes). A correlation coefficient of zero indicates no relationship between the variables. However, correlations are limited to linear relationships between variables. Even if the correlation coefficient is zero, a non-linear relationship might exist.

Fun Fact! *Correlation is a widely used term in statistics. In fact, it entered the English language in 1561, 200 years before most of the modern statistic tests were discovered. It is derived from the [same] Latin word correlation, which means relation.*

Bonferroni Correction: If one conducts a lot of correlations, some relationships will occur by chance. To mitigate this, Bonferroni correction is applied. It reduces the alpha level for the analysis, thus reducing the likelihood of making a Type I error (false positive); it is based on the number of times each variable is used.

Correlation Coefficient (*r*): Ranges from -1 to 1; describes to the strength of the relationship between the variables.

Critical Value: The minimum value at which an observed correlation coefficient is statistically significant.

***p*-value:** The probability of obtaining the observed results if the null hypothesis is true. A result is usually considered statistically significant if the *p*-value is ≤ .05.

Independent Samples *t*-Test

The independent samples *t*-test is used to determine if there is a significant difference between two groups (e.g., men vs. women) on a scale-level dependent variable. This test uses the difference between the average scores of the two groups to compute the *t* statistic, which is used with the *df* to compute the *p*-value (i.e., significance level). A significant result indicates the observed test statistic would be unlikely under the null hypothesis. The independent samples *t*-test carries the assumptions of independence of observations, normality, and equality (or homogeneity) of variance.

Independent Samples *t*-Test Formula:

$$t = \frac{\overline{X_1} - \overline{X_2}}{SE}$$

$\bar{X}_1$ = sample 1 mean
$\bar{X}_2$ = sample 2 mean
SE = standard error for sample 1 and sample 2 (using pooled-variance or separate-variances formula)

Fun Fact! *William Sealy Gosset, who published a paper about the t distribution in 1908, worked for the Guinness Brewery in Dublin, Ireland.*

Cohen's *d*: Effect size for the *t*-test; determines the strength of the differences between the matched scores. The larger the effect size, the greater the differences in the matched scores.

Degrees of Freedom (*df*): Refers to the number of values used to compute a statistic. The *df* is determined by the number of observations in the sample and equal the number of observations - 1; used with *t* to compute the *p*-value.

Levene's Test: Test to assess if the assumption of equality of variance is met; if significance is found, the groups differ in their spread of the dependent variable scores; this may differ from the output found from other statistical packages (such as SPSS), as Intellectus Statistics™ uses the median instead of the mean for calculations; the median tends to provide a more-robust choice that can account for non-normality.

Mean (*M*): The average value of a scale-level variable.

Normality: Refers to the distribution of the data. The assumption is that the data follows the bell-shaped curve.

***p*-value:** The probability of obtaining the observed results if the null hypothesis is true. A result is usually considered statistically significant if the *p*-value is ≤ .05.

Shapiro-Wilk Test: A test to assess if the assumption of normality is met. If statistical significance is found in this test, the data is *not* normally distributed.

Standard Deviation (*SD*): The spread of the data around the mean of a scale-level variable.

***t*-Test Statistic (*t*):** Used with the *df* to determine the *p* value.

Mann Whitney *U*

The Mann-Whitney *U* is a non-parametric test used to assess for significant differences in a scale or ordinal dependent variable by a single dichotomous independent variable. It is the non-parametric equivalent of the independent sample *t* -test. The test uses the mean ranks of the scores in each group to compute the *U* statistic, which in turn is used to compute the *p* -value (i.e., significance level). A significant result for this test suggests that the two groups have reliably different scores on the dependent variable. The Mann-Whitney *U* test assumes that the observations are independent of each other and that the dependent variable has a scale or ordinal level of measurement.

Fun Fact! *The Mann-Whitney U test is named after mathematician Henry Mann and his student Donald Whitney, who collaborated on the development of the test in the 1940s.*

Mean Rank: The average rank of the data for that group once the data is sorted and ranked.

Non-Parametric Test: A type of statistical test that does not require the data to follow a particular distribution; typically used when assumptions of a parametric test are violated or when the data do not fit the level of measurement required by a parametric test.

***p*-value:** The probability of obtaining the observed results if the null hypothesis (no relationship between the independent variable(s) and dependent variable) is true; in most social science research, a result is considered statistically significant if this value is ≤ .05.

***U*-Test Statistic (*U*):** Used to compute the *p* value.

Multiple Linear Regression

The multiple linear regression is the most common form of linear regression analysis. As a predictive analysis, the multiple linear regression is used to explain the relationship between one continuous dependent variable from two or more independent variables. It does this by creating a linear combination of all the independent variables to predict the dependent variable. The independent variables can be continuous or categorical (dummy coded as appropriate). The R^2 statistic is used to assess how well the regression predicted the dependent variable. While the unstandardized beta (B) describes the increase or decrease of the independent variable(s) with the dependent variable.

Fun Fact! *Linear regression and multiple linear regression are commonly used in economics and finance as a way to predict risk of investment, consumption spending, inventory investment, labor demand, and more!*

95% Confidence Interval (95% CI): An interval that estimates the range one would expect B to lie in 95% of the time given the samples tested comes from the same distribution.

Akaike's Information Criterion (*AIC*): A measure of model quality or fit. It uses the maximized log likelihood value as a baseline for model fit, and adds a penalty for estimating additional parameters. Smaller *AIC* values represent better model fit.

Degrees of Freedom (*df*): Used with the F ratio to determine the p-value.

Dummy-Code: Performed in order to add a nominal or ordinal independent variable into the regression model; turns the one variable into a series of dichotomous "yes/no" variables, one for each category; one of the categories are left out of the regression as the reference group that all other categories are compared to.

***F* Ratio (*F*):** Used with the two *df* values to determine the p value of the overall model.

Homoscedasticity: Refers to the relationship between the residuals and the fitted values; the assumption is met when the residuals plot has the points randomly distributed (with no pattern), and the distribution line is approximately straight.

Multicollinearity: A state of very high intercorrelations or inter-associations among a set of variables.

Normality: Refers to the distribution of the residuals; the assumption is that the residuals follow a bell-shaped curve; the assumption is met when the q-q plot has the points distributed approximately on the normality line.

Outlier: A data point that is abnormally distant from a set of observations.

***p*-value:** The probability that the null hypothesis (no relationship in the dependent variable by the independent variable) is true.

Residuals: Refers to the difference between the predicted value for the dependent variable and the actual value of the dependent variable.

R-Squared Statistic (R^2): Tells how much variance in the dependent variable is explained by only the predictor variables.

Standardized Beta (*β*): Ranges from -1 to 1; gives the strength of the relationship between the predictor and dependent variable.

Studentized Residuals: Residuals that are scaled by diving the each residual by the estimated standard deviation of the residuals.

***t*-Test Statistic (*t*):** Used with the *df* to determine the *p* value; also can show the direction of the relationship between the predictor and dependent variable.

Unstandardized Beta (*B*): The slope of the predictor with the dependent variable.

Standard Error (*SE*): How much the *B* is expected to vary.

Variance Inflation Factors: A measurement to assess the amount of multicollinearity present in regression analysis.

Raw Output

Descriptives

Included Variables:
Gender, DP, EE, and PA

Sample Size (Complete Cases):
N = 15

Summary Statistics: Frequency Table for Nominal Variables

Variable	n	%
Gender		
Male	7	46.667
Female	8	53.333

Summary Statistics: Scale

Variable	M	SD	n	95% CI	SE_M	Min	Max	Mdn	Mode
DP	49.067	32.217	15	[31.226, 66.908]	8.318	12.000	100.000	38.000	38.000
EE	46.733	30.125	15	[30.051, 63.416]	7.778	12.000	99.000	38.000	79.000
PA	48.533	31.202	15	[31.254, 65.812]	8.056	12.000	99.000	37.000	42.000

Quantiles:

	DP	EE	PA

Min	12.000	12.000	12.000
10%	16.800	15.200	17.600
20%	20.400	19.400	20.800
25%	22.500	22.000	22.500
30%	24.800	24.800	24.800
40%	29.800	29.200	28.600
50%	38.000	38.000	37.000
60%	50.400	48.800	51.600
70%	66.800	65.200	69.200
75%	77.500	72.500	75.000
80%	88.200	79.000	82.200
90%	94.400	87.400	92.200
Max	100.000	99.000	99.000

Pearson Correlation Test

Included Variables:
DP, EE, and PA

Sample Size (Complete Cases):
N = 15

Correlation Matrix:

Variable	1	2	3
1. DP	-		
2. EE	0.966*	-	
3. PA	0.980*	0.987*	-

Note. '*' indicates p < 0.0500.

Correlation Results:

Combination	r	95.000% CI	n	p
DP-EE	0.966	[0.899, 0.989]	15	4.747×10^{-09}
DP-PA	0.980	[0.939, 0.994]	15	3.405×10^{-10}
EE-PA	0.987	[0.960, 0.996]	15	3.029×10^{-11}

Note: p-values adjusted using the Holm correction.

Independent t-Test for DP by Gender

Included Variables:
DP and Gender

Sample Size (Complete Cases):
N = 15

Shapiro-Wilk Test:
Male: W = 0.884, p = 0.244
Female: W = 0.863, p = 0.129
Overall: W = 0.868, p = 0.0312

Levene's Test:
$df_n = 1$, $df_d = 13$, F = 4.963, p = 0.0442

Results:

	Male			Female					
Variable	M	SD	n	M	SD	n	t	p	d
DP	24.143	10.839	7	70.875	28.468	8	-4.301	0.00188	2.170

Note. n = 15, df = 9.216.

Confidence Interval Based on α = 0.0500:
Lower Limit = -71.226, Mean Difference = -46.732, Upper Limit = -22.238

Two-Tailed Mann Whitney U Test for DP by Gender

Included Variables:
DP and Gender

Sample Size (Complete Cases):
N = 15

Results:
U = 4.000, z = -2.900, p = 0.00373

Medians for DP by Gender
Male (n = 7) = 21.000 and Female (n = 8) = 77.500

Linear Regression with DP predicted by EE and PA

Included Variables:
DP, EE, and PA

Sample Size (Complete Cases):
N = 15

Lilliefors-Kolmogorov-Smirnov Test:
D = 0.222, p = 0.0445

Variance Inflation Factors:

Variable	VIF
EE	38.937

PA	38.937

Durbin-Watson Test:
DW = 2.308, p = 0.662

Linear Regression Coefficients:

Variable	B	SE	95.00% CI	β	t	p
(Intercept)	-0.0249	3.394	[-7.419, 7.369]	0.000	-0.00733	0.994
EE	-0.0368	0.383	[-0.872, 0.798]	-0.0344	-0.0959	0.925
PA	1.047	0.370	[0.241, 1.853]	1.014	2.829	0.0152

Model Fit Statistics:
$F(2,12) = 145.529$, $p = 3.854 \times 10^{-09}$, $R^2 = 0.960$, adj. $R^2 = 0.954$

Partial Correlations:

Variable	Estimate	t	p
EE	-0.0277	-0.0959	0.925
PA	0.632	2.829	0.0152

Semi-Partial Correlations:

Variable	Estimate	t	p
EE	-0.00551	-0.0191	0.985
PA	0.162	0.570	0.579