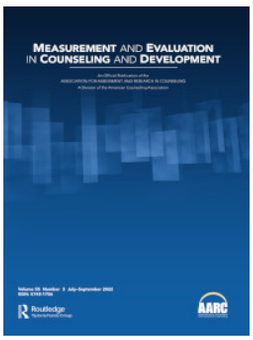

Routledge
Taylor & Francis Group

Measurement and Evaluation in Counseling and Development

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/uecd20

Screening Utility of the PHQ-2 and PHQ-9 for Depression in College Students: Relationships with Substantive Scales of the MMPI-3

Nicole M. Morris, Paul B. Ingram, Sean M. Mitchell & Sarah E. Victor

To cite this article: Nicole M. Morris, Paul B. Ingram, Sean M. Mitchell & Sarah E. Victor (2022): Screening Utility of the PHQ-2 and PHQ-9 for Depression in College Students: Relationships with Substantive Scales of the MMPI-3, Measurement and Evaluation in Counseling and Development, DOI: 10.1080/07481756.2022.2110899

To link to this article: https://doi.org/10.1080/07481756.2022.2110899

MEASUREMENT AND EVALUATION IN COUNSELING AND DEVELOPMENT
https://doi.org/10.1080/07481756.2022.2110899

Routledge
Taylor & Francis Group

ASSESSMENT, DEVELOPMENT, AND VALIDATION

Screening Utility of the PHQ-2 and PHQ-9 for Depression in College Students: Relationships with Substantive Scales of the MMPI-3

Nicole M. Morris[a], Paul B. Ingram[a,b], Sean M. Mitchell[a], and Sarah E. Victor[a]

[a]Texas Tech University, Lubbock, TX, USA; [b]Eastern Kansas Veteran Healthcare System, Topeka, KS, USA

ABSTRACT

We investigated the validity and screening effectiveness of the PHQ-2 and PHQ-9 scores in 229 college students in a cross-sectional design. PHQ associations with Minnesota Multiphasic Personality Inventory-3 internalizing scales suggest PHQ scores are effective screening tools for college students and may aid in effective triage and service needs.

KEYWORDS

MMPI-3; college health; psychological assessment; depression

The Healthy Minds Network (2021) estimates that around 21% of university students have screened positive for a major depressive episode in the past year, a 3% increase from 2019 estimates obtained prior to the COVID-19 pandemic (Healthy Minds Network, 2019). Similarly, estimates suggest that around 12% of U.S. college students are at risk of suicide (Kim et al., 2020), underscoring the substantial effect of depressive experiences on day-to-day life. Given that depression is so common and impactful for college students (Auerbach et al., 2018; Liu et al., 2019), and counseling is effective for alleviating some of these concerns (Cuijpers et al., 2016), early identification and monitoring of these depressive symptoms offers an opportunity to improve clinical care. Such efforts may play a critical role in addressing the impact of depression, as well as the substantial burden it places on the economic and educational systems in which students operate (e.g., Eisenberg et al., 2007; Santomauro et al., 2021).

The Patient Health Questionnaire (PHQ-9; Spitzer et al., 1999) is a commonly used nine-item depression screening instrument that was developed using the Primary Care Evaluation of Mental Disorders (PRIME-MD), which utilized criteria from the Diagnostic and Statistical Manual of Mental Disorders 4th Edition (DSM-IV; American Psychiatric Association [APA], 1994), as a guide to categorical classification of depressive disorders. The PHQ-9 is incorporated into medical care systems around the world. The PHQ-9, as well as the more abbreviated two-item PHQ-2, has been widely used in evidence-based practices to monitor symptom change across a variety of medical treatment settings (e.g., Veteran Affairs Medical Care System or Primary Care Setting; Ingram et al., 2021). It is also frequently used as a screener to determine if more intensive psychological assessment and treatment planning are needed (Maurer, 2012).

Research has regularly supported the use of PHQ instruments' (PHQ-9 and PHQ-2) with high sensitivity and specificity estimates (.80+) and increased risk for major depression across various cut-scores (Moriarty et al., 2015). However, some studies have also warned that PHQ instruments are not effective in assessing intensity of symptoms and may not effectively capture all important elements of depressogenic symptoms necessary for diagnostic certainty (Mitchell et al., 2016), particularly those of low mood (i.e., Major Depressive Episode Criteria A1; see

CONTACT Nicole M. Morris nmlemaster1@gmail.com Department of Psychological Sciences, Texas Tech University, 2810 18th Street, Lubbock, TX 79424, USA.

Supplemental data for this article is available online at https://doi.org/10.1080/07481756.2022.2110899.

American Psychiatric Association (APA), 2022; Malpass et al., 2016). As a result of these concerns and general screening purposes, sole use of the PHQ-9 to determine depression diagnoses may overestimate actual prevalence rates of depression, warranting reconsideration of cut-score effectiveness (Levis et al., 2020). Similarly, studies on the PHQ-2 have found its diagnostic accuracy to be lower than initially reported (Manea et al., 2016). In short, the PHQ offers a validated framework with several psychometric strengths; however, it also faces some limitations as a screening platform that has substantial decision-making implications within medical settings. Regardless of these limitations, the PHQ instruments are widely used patient outcome measures with potential to increase identification of depressive symptoms and treatment recommendation for healthcare providers across college health centers.

To expand knowledge about the construct validity of the PHQ-2 and PHQ-9 scale scores within the framework of contemporary conceptualizations of psychopathology (see Kotov et al., 2017), McCord and Provost (2020) examined the relationship of Minnesota Multiphasic Personality Inventory-2-Restructured Form (MMPI-2-RF; Tellegen & Ben-Porath, 2008/2011) scale scores to PHQ-2 and PHQ-9 total scores. Consistent with depressive presentations, PHQ-9 scale scores were strongly associated with somatic and cognitive complaints as well as internalizing difficulties, specifically those related to negative affectivity. However, McCord and Provost (2020) also found that the MMPI-2-RF scale assessing anhedonia did not meet *a priori* thresholds for a meaningful relationship. As such, they concluded that anhedonic experiences (measured on Restructured Clinical Scale 2 [RC2] of the MMPI-2-RF) are not effectively assessed by the PHQ-9 and suggest interpreting PHQ-9 scores as indicative of general emotional distress rather than depression per se. They reported similar patterns of performance for PHQ-2 scores and concluded that the total score calculated with either PHQ instrument may inappropriately combine two distinct diagnostic aspects of depressive disorders (low mood [Criteria A1] and anhedonia [Criteria A2]; American Psychiatric Association [APA], 2022) without accounting for important differences in their presentation. Exemplifying ways in which this combination is problematic, McCord and Provost (2020) point out that elevated PHQ scores may not assist in making determinations for specific anti-depressant medications since serotonin and dopamine pathways play distinct roles in these symptom criteria (see Nutt, 2008).

Although McCord and Provost's (2020) study offered some important directions in understanding the construct validity of the PHQ-9 and PHQ-2, it was also limited in some important ways. First, scale scores and severity level frequency of PHQ instruments were not reported. This missing descriptive information has substantial interpretive implications since their sample was recruited solely from an undergraduate participant pool, meaning that scores were likely lower than might be seen in even college samples with bona-fide depressive symptoms. Likewise, it is impossible to know if the reported associations accurately reflect those seen in individuals screening positively on PHQ instruments. Additionally, their results capture only correlations between PHQ instruments total score with MMPI-2-RF criteria, leaving the PHQ item-level associations unassessed. In many ways, this narrower interpretive information may be just as critical to decision making about recommendations given that PHQ items are available to physicians and generally align to diagnostic criteria for a major depressive episode (Spitzer et al., 1999).

Current Study

Expanding interpretive utility of the PHQ instruments using a latent broadband measure of personality and psychopathology (i.e., MMPI-3; Ben-Porath & Tellegen, 2020a, 2020b) allows us to expand diagnostic screening practices through contemporary models of psychopathology. Expansion of PHQ instruments construct validity amongst college students can aid in the interpretation and treatment recommendations made by healthcare providers working with college students and screening for depression. Therefore, in this study, we will extend the findings of McCord and Provost (2020). To aid interpretations and treatment recommendations made by healthcare providers utilizing PHQ screening instruments, we investigated the screening efficiency

and PHQ-2 and PHQ-9 scale score, item-level, and high score relationships with the internalizing substantive scales of the MMPI-3.

Method

Participants

This study drew its sample from two groups of participants ($N = 229$), both composed of college students at a large, public university in the southwestern United States. Subsamples were combined for this study (see Table 1 for demographic descriptive of each sample). In general, participants in the final sample identified primarily as White (54.6%), women (79.0%), and their ages ranged from 18 to 56 years ($M = 20.2$, $SD = 3.8$). One set of participants ($n = 83$) was recruited via a psychology department subject pool in exchange for course credit and represents a more general college sample, as they were not screened for depressive symptoms. The second set of participants ($n = 146$) was recruited to ensure a full range of depression symptoms within the sample and enable clinically meaningful interpretations from the data. This sample was recruited through flyers posted throughout common spaces of the university (e.g., gym or student union), as well as through email advertisements, and participants were paid $20 for their participation. For inclusion in the depressive subsample, individuals had to screen positively for at least moderate depressive symptoms (score ≥ 10) on the PHQ-9, corresponding to commonly used screening practices.

Instruments

MMPI-3

The MMPI-3 (Ben-Porath & Tellegen, 2020a, 2020b) is a broadband measure of psychopathology and personality consisting of 335 true-false items. The MMPI-3 consists of 10 validity scales and 42 substantive scales which hierarchically measure psychopathology in line with contemporary models of psychopathology (i.e., Kotov et al., 2017). The MMPI-3 normative sample was updated to reflect the 2020 United States census data. Most relevant to the current study are those scales that assess broad internalizing difficulties (i.e., Emotional/Internalizing [EID]), demoralization (i.e., RCd), low positive emotions (i.e., RC2), dysfunctional negative emotions (i.e., RC7), suicidal/death ideation (i.e., SUI), helplessness/hopelessness (i.e., HLP), self-doubt (i.e., SFD), inefficacy (i.e., NFC), stress (i.e., STR), worry (i.e., WRY), compulsivity (i.e., CMP), anxiety-related experiences (i.e., ARX), anger proneness (i.e., ANP), behavior-restricting fears (i.e., BRF), high levels of negative emotions (NEGE), and lack of positive emotions and social avoidance (INTR). MMPI-3 scales are

Table 1. Demographic Characteristics for Combined Samples ($N = 229$).

	Depressed Sub-Sample ($n = 146$)	General Student Sub-Sample ($n = 83$)
	N (%)	*N (%)*
Gender		
Men	37 (25.3%)	11 (13.3%)
Women	109 (74.7%)	72 (86.7%)
Ethnicity		
Asian	6 (4.1%)	2 (2.4%)
Black/African American	12 (8.2%)	8 (9.6%)
Latinx	47 (32.2%)	25 (30.1%)
Multiracial	–	2 (2.4%)
Other	-	2 (2.4%)
White	81 (55.5%)	44 (53.0%)
	M (SD)	*M (SD)*
Age	20.0 (1.8)	20.6 (5.8)
PHQ-9 Total Score	15.2 (5.2)	5.7 (5.4)
PHQ-2 Total Score	3.5 (1.6)	1.2 (1.5)

Note. N = sample size; M = mean; SD = standard deviation; Differences in the distribution of gender across samples were tested and indicated significant and negligible differences, X^2 (1, $n = 229$) = 4.67, p = .03, Cramer's v = .14.

interpreted as uniform *T* scores in which 50 is average. Substantive scales are typically interpreted as clinically elevated if scale scores meet or exceed *T*65, corresponding to the 92nd percentile in the normative sample (see Ben-Porath & Tellegen, 2020a). Likewise, MMPI-3 substantive scale *T* scores of 75 or greater corresponds to the 99th percentile and is a another commonly used cut-score within MMPI research. Extensive information on the validation and initial support for the MMPI–3's score reliability and validity within a census-matched normative sample ($n = 1{,}620$), and with use in college student specialty comparison groups specifically, are reported in the *MMPI–3 Technical Manual* (Ben-Porath & Tellegen, 2020b). See Table 2 for internal reliability estimates for each of the MMPI-3 scales. Generally, scale scores demonstrated strong reliability with an average of .80 and a range of .42 to .93. Consistent with findings in the *MMPI-3 Technical Manual* (Ben-Porath & Tellegen, 2020b), BRF scale scores produced low internal reliability (α = .42).

PHQ-9

The PHQ-9 is a 9-item self-report inventory of the frequency of depressive symptoms during the previous 2 weeks using a 4-point response scale (0 = *not at all*, 1 = *several days*, 2 = *more than half the days*, 3 = *nearly every day*). Items of the PHQ-9 and PHQ-2 assess the diagnostic criteria of a major depressive episode. Research supports the psychometric properties of the PHQ-9 and PHQ-2 with higher scores indicating increased depressive experiences, particularly negative mood (Boothroyd et al., 2019; Kroenke et al., 2001; Kroenke & Spitzer, 2002). A single cut score of greater than 9 is frequently used in clinical practice to screen for the possible presence of depressive symptoms (see Kroenke & Spitzer, 2002). In this study, the PHQ-9 had a coefficient alpha of .90 and a mean of 11.8 (*SD* = 7.0). Across groups, 89 participants screened negative for possible depression on the PHQ-9 and 140 screened positively.

PHQ-2

PHQ-2 is an abbreviated version of the PHQ-9 consisting of items 1 and 2 of the original PHQ-9. The PHQ-2 was designed to serve as a very brief measure of depressive symptoms for use in extremely busy clinical settings in which depression screenings are needed (Kroenke et al.,

Table 2. *T*-Test Results for All MMPI-3 Internalizing Scales Across PHQ-9 Classification Groups ($N = 229$).

		Screened Negative for Depression (PHQ-9 score < 10) $n = 89$			Screened Positive for Depression (PHQ-9 > 10) $n = 140$				
Scale	α	*M*	*SD*	%>*T*65	*M*	*SD*	%>*T*65	*t*-value	*g*
EID	.93	52.9	8.4	19.80%	70.5	10.2	77.80%	−13.51*	**1.83**
RCd	.91	51.5	8.8	20.70%	69.2	9.2	79.40%	−14.62*	**1.96**
RC2	.83	50.6	9.7	17.00%	63.5	13.2	59.00%	−8.48*	**1.07**
RC7	.85	54.6	10.2	29.60%	68.9	10.2	70.00%	−10.33*	**1.40**
SUI	.82	47.2	8.0	17.80%	61.0	16.8	43.60%	−8.32*	**0.98**
HLP	.80	47.4	9.4	14.30%	59.6	13.6	43.60%	−8.01*	**1.00**
SFD	.84	51.4	8.5	15.20%	67.8	9.9	71.80%	−13.33*	**1.74**
NFC	.80	53.2	9.3	24.10%	65.6	10.4	63.20%	−9.45*	**1.25**
STR	.63	56.1	10.9	33.10%	65.8	8.8	63.20%	−13.51*	**1.00**
WRY	.79	52.2	9.8	40.10%	66.4	7.9	75.20%	−12.01*	**1.62**
CMP	.79	55.3	11.1	37.50%	64.2	11.0	63.30%	−5.93*	**0.80**
ARX	.89	53.0	11.0	25.10%	70.0	12.3	62.40%	−10.87*	**1.43**
ANP	.87	51.9	9.3	12.3%	57.8	10.2	27.9%	−4.47*	**0.60**
BRF	.42	55.4	11.8	17.10%	64.7	14.8	32.50%	−5.24*	**0.67**
NEGE-r	.88	54.5	10.3	25.90%	67.9	8.7	59.80%	−10.57*	**1.43**
INTR-r	.86	50.5	10.1	17.10%	60.9	13.6	49.50%	−6.57*	**0.83**

Note. α = Cronbach's alpha; *t*-value = *t*-test value; *M* = Mean; *SD* = Standard Deviation; %>*T*65 = percent of individuals exceeding a cut-score of *T*65; *p-value < .003 (Bonferroni corrected p-value = .05/15 = .003); **bold** = effect size difference was significant ($p < .003$); EID = Emotional/Internalizing Dysfunction; RCd = Demoralization; RC2 = Low Positive Emotions; RC7 = Dysfunctional Negative Emotions; SUI = Suicidality; HLP = Helplessness/hopelessness; SFD = Self-doubt; NFC = Inefficacy; STR = Stress; WRY = Worry; CMP = Compulsivity; ARX = Anxiety Related Experiences; ANP = Anger Proneness; BRF = Behavior Restricting Fears; NEGE-r = Negative Emotionality/Neuroticism; INTR-r = Introversion/Low Positive Emotionality; *g* = Hedge's g effect size.

2003). The PHQ-2 maintains the same 4-point response scale and scoring methods as the PHQ-9, with items assessing low positive emotions (i.e., item 1) and feelings of depression and hopelessness (i.e., item 2). PHQ-2 total score ranges from 0 to 6 with a cut-score of 3 or greater indicating the likely presence of depressive symptoms. In this study, PHQ-2 scale scores were derived from participants responses to the PHQ-9, produced a coefficient alpha of .84, a mean of 2.7 ($SD = 1.9$), and similar distribution of participants across groups ($n = 112$ [screened negative for depression] and $n = 117$ [screened positive for depression]).

Procedures

This study received IRB approval. A total of 243 participants completed the MMPI-3 and collateral measures. Collateral measures varied slightly across subsamples as those in the subsample screened for depression received measures predominantly assessing internalizing symptoms and those in the general college student subsample receiving measures of eating pathology symptoms. Participants were excluded from data analyses if they did not produce a valid MMPI-3 profile. Potential participants were excluded based on standard interpretive recommendations on MMPI-3 non-content based (Cannot Say [CNS] > 18, or Combined Response Inconsistency [CRIN], Variable Response Inconsistency [VRIN], or True Response Inconsistency [TRIN] ≥ 80*T*) and content-based over-reporting (F ≥ 100, and Fp ≥ 100) scales (Ben-Porath & Tellegen, 2020a, 2020b). After exclusions based on invalid MMPI-3 profiles, the final sample was 229 participants (excluded $n = 14$ [general college student subsample = 6; college students screened for depression subsample = 8]).

Participants were classified into groups based on their PHQ-9 and PHQ-2 total scores[1]. Those that screened positive for depression (PHQ-9: ≥ 10; PHQ-2 ≥ 3; (see Kroenke et al., 2001) and those that did not. Groups scale scores were compared across MMPI-3 internalizing substantive scale scores using a series of *t*-tests. Effect sizes were calculated between groups using *Hedges' g,* which provides a sample size adjusted estimate of effect magnitude with values greater than |.4| indicating clinically meaningful differences (Ferguson, 2009), corresponding to an adjusted statistical significance of $p \leq .003$ (alpha = .05/15 correlations. Additionally, zero-order bivariate correlations were calculated between MMPI-3 internalizing scale scores and PHQ-2 and PHQ-9 total scores and individual items. We interpreted the magnitude of these relationships using recommended ranges of (Cohen, 1988): small ($.1 < r < .3$), medium ($.3 < r < .5$), and large ($r \geq .5$). To account for shard method variance, we also interpreted effects as clinically meaningful only if they demonstrated at least a medium effect ($r > .3$), corresponding to an adjusted statistical significance of $p \leq .003$ (alpha = .05/15 correlations). Finally, relative risk ratios (RRR) were computed to identify the increased risk of internalizing symptoms for those screened for probable depression using PHQ-9 and PHQ-2 cut scores (PHQ-9 ≥ 10; PHQ-2 ≥ 3). Four cut-scores on the MMPI-3 internalizing scales were investigated: 60 *T*, 65 *T*, 70 *T*, and 75 *T*. RRR are considered meaningful if the lower and upper bound of the 95% confidence intervals do not cross 1. When the lower/upper bounds cross a score of 1 (i.e., a RRR of 1.17 with a lower bound of .98 and an upper bound of 1.25) it indicates that there is not sufficient confidence to ensure that the observed RRR differs from the null (i.e., RRR of 1.0).

Results

PHQ-9

Two severity groups (those screening negative for depressive symptoms [score < 9; $n = 89$] and those screening positively for depressive symptoms [score ≥ 10; $n = 140$]) were created using recommended PHQ-9 cut-scores. We compared groups across all of the MMPI-3 internalizing scales (i.e., EID, RCd, RC2, RC7, SUI, HLP, SFD, NFC, STR, WRY, CMP, ARX, ANP, BRF, NEGE-r, and INTR-r). Results indicated statistically significant differences across all scales (Table 2). Large, significant effects were observed across all internalizing MMPI-3 scales, except for BRF and ANP.

Largest effect size differences were observed across scales assessing demoralization, general internalizing symptoms, and feelings of self-doubt and worry ($g = 1.62$ to 1.96).

Next, RRR were computed across four different cut-scores (i.e., 60 *T*, 65 *T*, 70 *T*, and 75 *T*) to investigate the risk of an elevated MMPI-3 symptom scale depending on depression screening status using the PHQ-9 total score (negative depression screening < 9 on PHQ-9 [$n = 89$]; positive depression screening ≥ 10 on PHQ-9 [$n = 140$]; Table 3). All RRR using the PHQ-9 indicated meaningful risk between those screening positive and negative for depression across MMPI-3 cut-scores. Broadly, RRR analyses suggest that individuals with higher MMPI-3 scale scores have the highest rate of likely depression based on the cut-scores provided for the PHQ-9. At the MMPI-3 recommended cut value of 65 *T*, RRR indicate all internalizing scales are associated with significantly increased risk of PHQ-9 elevation (1.91 [CMP] to 12.2 [SFD]). Average RRR increased as MMPI-3 scale cut-scores increased, from 60 *T* ($M_{RRR} = 3.79$) to 75 *T* ($M_{RRR} = 10.95$). Scales assessing general internalizing symptoms (EID), demoralization (RCd), and self-doubt (SFD) were consistently amongst the highest observed RRR across each cut-score level.

PHQ-2

Two severity groups (those screening negatively for depressive systems [score < 3; $n = 112$] and those screening positively for depression [score ≥ 3; $n = 117$]) were created using recommended PHQ-2 cut-scores, and we compared these groups with the same methods used for the PHQ-9. Significant group differences were observed across all of the examined internalizing scales, with the exception of scores on BRF (Table 4). In general, effects were mostly large (87%; $g = .83$ to 1.63) with the remainder falling within a medium effect size range ($g = .59$ to .76), consistent with the patterns observed on the PHQ-9. Further, RRR results for the PHQ-2 (score ≥ 3; Table 5) suggest that individuals with MMPI-3 scale scores above 65 *T* for all the internalizing scales were associated with higher odds of screening positive for depression on the PHQ-2. As with the PHQ-9, RRR increased for the MMPI-3 cut-score across 60 *T* (1.53 [WRY] to 4.74 [SFD]), 65 *T* (1.59 [ANP] to 4.74 [SFD]), 70 *T* (1.69 [ANP] to 13.89 [HLP]), and 75 *T* (1.52 [BRF] to 13.89 [HLP]).

PHQ Item Content Relationship

Correlations were also computed between PHQ-2 and PHQ-9 items and total scores with each of the MMPI-3 internalizing scales (Table 6). PHQ-2 and PHQ-9 total scores demonstrated

Table 3. Relative Risk Ratios (RRR) for PHQ-9 (cut-score > 10) Across MMPI-3 Internalizing Scales ($N = 229$).

Scale	RRR at 60*T* Cut-point	RRR 95% CI	RRR at 65*T* Cut-point	RRR 95% CI	RRR at 70*T* Cut-point	RRR 95% CI	RRR at 75*T* Cut-point	RRR 95% CI
EID	**5.59**	3.36–9.26	**7.35**	3.92–13.70	**9.80**	4.12–23.26	**15.63**	3.89–62.5
RCd	**5.46**	3.29–9.09	**5.50**	3.23–9.43	**10.53**	4.44–25.00	**29.41**	4.10–2.00
RC2	**2.94**	1.94–4.48	**4.03**	2.33–6.94	**9.35**	2.99–29.41	**10.53**	2.58–43.48
RC7	**2.75**	1.95–3.88	**3.94**	2.49–6.21	**4.95**	2.60–9.35	**6.76**	2.80–16.13
SUI	**3.47**	1.93–6.21	**3.47**	1.93–6.21	**6.67**	2.48–17.86	–	–
HLP	**4.10**	2.14–7.87	**4.10**	2.14–7.87	**9.26**	2.26–37.04	**9.26**	2.26–37.04
SFD	**12.20**	5.18–28.57	**12.20**	5.18–28.57	**9.90**	4.18–23.81	**31.25**	4.39–2.00
NFC	**2.85**	1.98–4.10	**4.31**	2.56–7.25	**6.10**	2.92–12.66	**5.88**	2.17–15.87
STR	**2.19**	1.53–3.13	**2.19**	1.53–3.13	**2.90**	1.48–5.65	**2.90**	1.48–5.65
WRY	**3.60**	2.42–5.32	**3.60**	2.42–5.32	**6.37**	3.24–12.50	–	–
CMP	**1.91**	1.38–2.64	**1.91**	1.38–2.64	**4.65**	2.07–10.53	**4.65**	2.07–10.53
ARX	**4.20**	2.61–6.67	**4.31**	2.56–7.25	**4.63**	2.61–8.20	**8.40**	3.15–22.22
ANP	**1.64**	1.36–2.61	**1.56**	1.18–1.76	**1.61**	1.14–1.82	–	–
BRF	**1.89**	1.36–2.61	**2.99**	1.59–5.62	**2.99**	1.59–5.62	**4.29**	1.55–11.90
NEGE-r	**3.13**	2.20–4.42	**3.56**	2.20–5.75	**3.39**	1.95–5.92	**5.08**	2.09–12.35
INTR-r	**2.61**	1.68–4.07	**3.45**	1.98–5.99	**6.37**	2.35–17.24	**8.26**	2.01–34.48

Note. **Bold** = RRR values differ meaningfully in risk between those screening positive and negative for depression based on the PHQ-9; EID = Emotional/Internalizing Dysfunction; RCd = Demoralization; RC2 = Low Positive Emotions; RC7 = Dysfunctional Negative Emotions; SUI = Suicidality; HLP = Helplessness/hopelessness; SFD = Self-doubt; NFC = Inefficacy; STR = Stress; WRY = Worry; CMP = Compulsivity; ANP = Anger Proneness; ARX = Anxiety Related Experiences; BRF = Behavior Restricting Fears; NEGE-r = Negative Emotionality/ Neuroticism; INTR-r = Introversion/Low Positive Emotionality; When a RRR has a confidence interval that crosses 1, this indicates no reliable difference in risk.

Table 4. *T*-Test Results for all MMPI-3 Internalizing Scales Across PHQ-2 Classification Groups ($N = 229$).

	Screened Negative for Depression (PHQ-2 score < 3) $n = 112$			Screened Positive for Depression (PHQ-2 > 3) $n = 117$				
Scale	*M*	*SD*	%>*T*65	*M*	*SD*	%>*T*65	*t*-value	*g*
EID	55.6	9.6	19.8%	71.4	10.6	77.8%	−11.79*	**1.55**
RCd	54.2	10.6	20.7%	70.0	8.8	79.4%	−12.32*	**1.62**
RC2	52.2	10.5	17.0%	64.5	13.4	59.0%	−7.67*	**1.01**
RC7	57.9	11.8	29.6%	68.5	10.4	70.0%	−7.25*	**.96**
SUI	49.4	10.4	17.8%	61.6	17.2	43.6%	−6.45*	**.85**
HLP	49.4	9.9	14.3%	60.1	14.4	43.6%	−6.56*	**.86**
SFD	54.1	10.3	15.2%	68.5	9.7	71.8%	−10.81*	**1.42**
NFC	56.2	10.5	24.1%	65.3	10.9	63.2%	−6.42*	**.85**
STR	58.1	11.1	33.1%	65.7	8.9	63.2%	−5.74*	**.76**
WRY	56	11.2	40.1%	65.5	8.8	75.2%	−7.14*	**.94**
CMP	57.3	11.4	37.5%	64.0	11.3	63.3%	−4.43*	**.58**
ARX	56.5	12.7	25.1%	70.0	12.9	62.4%	−7.97*	**1.05**
ANP	52.5	9.0	12.6%	58.4	10.6	30.8%	−4.59*	**.61**
BRF	58.2	13.5	17.1%	63.9	14.8	32.5%	−3.05	.40
NEGE-r	57.6	11.1	25.9%	67.6	9.4	59.8%	−7.35*	**.98**
INTR-r	51.6	10.6	17.1%	61.9	13.8	49.5%	−6.25*	**.82**

Note. *t*-value = *t*-test value; *M* = Mean; *SD* = Standard Deviation; %>*T*65 = percent of individuals exceeding a cut-score of *T*65; *p-value < .003 (Bonferroni corrected p-value = .05/15 = .003); **bold** = effect size different was significant ($p < .003$); EID = Emotional/Internalizing Dysfunction; RCd = Demoralization; RC2 = Low Positive Emotions; RC7 = Dysfunctional Negative Emotions; SUI = Suicidality; HLP = Helplessness/hopelessness; SFD = Self-doubt; NFC = Inefficacy; STR = Stress; WRY = Worry; CMP = Compulsivity; ARX = Anxiety Related Experiences; ANP = Anger Proneness; BRF = Behavior Restricting Fears; NEGE-r = Negative Emotionality/ Neuroticism; INTR-r = Introversion/Low Positive Emotionality; *g* = Hedge's g effect size; *d* = Cohen's *d* effect size.

Table 5. Relative Risk Ratios (RRR) for PHQ-2 (cut-score > 3) Across MMPI-3 Internalizing Scales ($N = 229$).

	RRR at 60*T*		RRR at 65*T*		RRR at 70*T*		RRR at 75*T*	
Scale	Cut-point	RRR 95% CI	Cut-point	RRR 95% CI	Cut-point	RRR 95% CI	Cut-point	RRR 95% CI
EID	**3.24**	2.34–4.46	**3.95**	3.06–5.81	**4.65**	2.79–7.75	**11.24**	4.18–30.30
RDd	**3.65**	2.58–5.16	**3.88**	2.66–5.65	**4.31**	2.67–6.94	**6.54**	2.89–14.71
RC2	**2.83**	2.45–5.38	**3.47**	2.25–5.38	**5.46**	2.56–11.77	**7.41**	2.70–20.41
RC7	**1.81**	1.42–2.32	**2.38**	1.75–3.25	**2.65**	1.73–4.05	**3.31**	1.89–5.81
SUI	**2.44**	1.56–3.82	**2.44**	1.56–3.82	**3.05**	1.63–5.68	**8.26**	2.58–26.32
HLP	**3.05**	1.86–5.03	**3.05**	1.86–5.03	**13.89**	3.39–55.56	**13.89**	2.39–55.56
SFD	**4.74**	3.01–7.46	**4.74**	3.01–7.46	**4.35**	2.65–7.14	**7.04**	3.12–15.87
NFC	**1.78**	1.38–2.31	**2.62**	1.84–3.75	**3.47**	2.13–5.65	**2.61**	1.38–4.95
STR	**1.92**	1.42–2.58	**1.92**	1.42–2.58	**1.86**	1.10–3.15	**1.86**	1.10–3.15
WRY	**1.53**	1.46–2.40	**1.87**	1.46–2.40	**2.87**	1.91–4.31	–	–
CMP	**1.69**	1.28–2.22	**1.69**	1.28–2.22	**2.46**	1.41–4.31	**2.46**	1.41–4.31
ARX	**2.25**	1.66–3.06	**2.49**	1.76–3.55	**2.67**	1.82–3.94	**4.50**	2.39–8.48
ANP	**1.62**	1.38–1.77	**1.59**	1.29–1.77	**1.69**	1.35–1.85	–	–
BRF	**1.66**	1.26–2.20	**1.92**	1.18–3.12	**1.92**	1.18–3.12	1.52	.77–2.98
NEGE-r	**1.93**	1.53–2.45	**2.31**	1.63–3.27	**2.68**	1.73–4.17	**2.96**	1.58–5.56
INTR-r	**2.30**	1.60–3.32	**2.92**	1.87–4.59	**6.06**	2.67–13.70	**8.00**	2.48–25.64

Note. **Bold** = RRR values differ meaningfully in risk between those screening positive and negative for depression based on the PHQ-2; EID = Emotional/Internalizing Dysfunction; RCd = Demoralization; RC2 = Low Positive Emotions; RC7 = Dysfunctional Negative Emotions; SUI = Suicidality; HLP = Helplessness/hopelessness; SFD = Self-doubt; NFC = Inefficacy; STR = Stress; WRY = Worry; CMP = Compulsivity; ARX = Anxiety Related Experiences; ANP = Anger Proneness; BRF = Behavior Restricting Fears; NEGE-r = Negative Emotionality/ Neuroticism; INTR-r = Introversion/Low Positive Emotionality; When a RRR has a confidence interval that crosses 1, this indicates no reliable difference in risk.

medium to large associations with most of the internalizing scales. The strongest and most consistent associations with PHQ-9 item and total scores were EID, RCd, SFD, and WRY. PHQ-9 item 9 had the least frequent significant and meaningful associations; however, it had a large relationship with SUI ($r = .70$) as one would expect given PHQ item 9's content focus.

Discussion

This study investigated the screening efficiency of PHQ-2 and PHQ-9 across the identified severity groups based on established cut-score values (Kroenke et al., 2001; Kroenke & Spitzer, 2002) and examined how these groups relate to MMPI-3 internalizing scales among college

Table 6. Pearson Correlations Between All MMPI-3 Internalizing Scale Scores and PHQ Total and Items ($N = 229$).

	PHQ-9 Total	PHQ-2 Total	Item 1*	Item 2*	Item 3	Item 4	Item 5	Item 6	Item 7	Item 8	Item 9
EID	**.74**	**.76**	**.66**	**.68**	**.46**	**.56**	**.54**	**.62**	**.54**	**.42**	**.54**
RCd	**.76**	**.74**	**.69**	**.68**	**.46**	**.57**	**.56**	**.67**	**.55**	**.45**	**.51**
RC2	**.53**	**.55**	**.51**	**.50**	**.30**	**.40**	**.39**	**.39**	**.37**	.29	**.44**
RC7	**.57**	**.48**	**.44**	**.46**	**.41**	**.47**	**.43**	**.50**	**.46**	**.34**	**.35**
SUI	**.54**	**.50**	**.47**	**.46**	**.31**	**.33**	**.32**	**.45**	**.36**	**.31**	**.70**
HLP	**.51**	**.50**	**.48**	**.45**	.27	**.35**	**.33**	**.45**	**.36**	**.36**	**.40**
SFD	**.70**	**.66**	**.61**	**.62**	**.41**	**.53**	**.51**	**.68**	**.51**	**.40**	**.50**
NFC	**.54**	**.47**	**.46**	**.42**	**.32**	**.41**	**.37**	**.47**	**.45**	**.40**	**.38**
STR	**.43**	**.36**	.29	**.38**	**.35**	**.35**	**.31**	**.34**	**.36**	.25	.26
WRY	**.63**	**.54**	**.49**	**.51**	**.48**	**.55**	**.49**	**.53**	**.49**	**.38**	.28
CMP	**.36**	**.30**	.25	**.30**	.23	**.35**	**.30**	**.32**	.26	.23	.17
ARX	**.63**	**.53**	**.47**	**.52**	**.46**	**.48**	**.45**	**.54**	**.52**	**.40**	**.36**
ANP	**.37**	**.33**	.28	**.34**	.28	.28	**.30**	**.30**	.27	.19	.24
BRF	**.37**	.26	.22	.25	**.33**	**.33**	**.30**	**.30**	.27	**.31**	.18
NEGE-r	**.60**	**.52**	**.44**	**.52**	**.46**	**.46**	**.42**	**.51**	**.49**	**.35**	**.37**
INTR-r	**.44**	**.47**	**.45**	**.43**	.25	**.34**	**.31**	**.32**	**.33**	.21	**.38**

Note. * = Items make up the PHQ-2; **bold** = $p < .003$ and $r >$ medium effect; EID = Emotional/Internalizing Dysfunction; RCd = Demoralization; RC2 = Low Positive Emotions; RC7 = Dysfunctional Negative Emotions; SUI = Suicidality; HLP = Helplessness/hopelessness; SFD = Self-doubt; NFC = Inefficacy; STR = Stress; WRY = Worry; CMP = Compulsivity; ARX = Anxiety Related Experiences; ANP = Anger Proneness; BRF = Behavior Restricting Fears; NEGE-r = Negative Emotionality/ Neuroticism; INTR-r = Introversion/Low Positive Emotionality.

students. In general, results provide support for both the PHQ-2 and PHQ-9 as effective screening tools, whose scores are strongly associated with general internalizing pathology, and feelings of self-doubt. helplessness, and demoralization. Consistent and large effect size differences were observed across all MMPI-3 internalizing scale scores between those experiencing varying levels of clinically indicated depression on the PHQ-9 and PHQ-2. Results suggest that clinicians and healthcare providers utilizing the PHQ screening tool(s) in collegiate settings may be confident that when elevations occur, the presence of internalizing pathology warrants follow-up assessment and potential clinical care. Below we discuss (1) the specific areas of internalization measured on the MMPI-3 which are most associated with PHQ-2 and PHQ-9 scale scores and scale elevations and (2) confidence in screening practices using the PHQ-2 and PHQ-9. Both the PHQ-2 and the PHQ-9 demonstrate similar patterns of results and are discussed together.

Almost all types of internalizing pathology measured by the MMPI-3 are associated with PHQ elevations at standard interpretive cut-scores (i.e., ≥ 10 on PHQ-9; ≥ 3 on PHQ-2). PHQ-2 and PHQ-9 items, total scores, and cut-score elevations are most strongly associated with broad internalizing symptoms, self-doubt, and feelings of demoralization and helplessness. These associations, as well as associations across all internalizing scales, are generally large. When these findings are compared to previous research regarding MMPI-2-RF and PHQ-9 scale score association among college students, we observed similar associations between MMPI-3 internalizing scales and PHQ-2 and PHQ-9 scale scores (e.g., associations with EID and RCd). However, relationships are notably stronger with the recently released MMPI-3 (e.g., the difference between correlations from the MMPI-2-RF and MMPI-3 is .26 between PHQ-9 total and EID; McCord & Provost, 2020). These increases in association strength are likely a product of the MMPI-3's revision and refinement of scale content. Our study provides an expanded capacity for clinicians working with college students to interpret PHQ-2 and PHQ-9 scores within the framework of a highly validated personality measure. While McCord and Provost (2020) examined item-level relationships with MMPI-2-RF scale scores among college students, they did not evaluate cut-score effectiveness or PHQ screening group differentiation.

Limitations and Future Directions

The present study must be considered within the scope of its limitations. While results support the use of PHQ instruments as screeners of internalizing symptoms, we were not able to explore

the PHQ instruments' ability to predict future internalizing symptoms, as our study was cross-sectional. Future research may benefit from exploring the PHQ instruments' ability to predict future symptoms as a means of further understanding how to utilize PHQ instruments for screening and triage related purposes. Further, PHQ instruments are wildly utilized in settings beyond college students, such as primary care settings. Thus, replication of the present study in primary care settings is warranted to understand the generalizability of the present findings in other settings in which depression screenings are critical.

Implications for Counseling Practice and Research

Our results suggest that using PHQ tool(s) in screening for internalizing symptoms is effective for college students. Indeed, with their elevated risk of depression and mental health problems (Auerbach et al., 2018; Ibrahim et al., 2013), it may be warranted and effective for colleges to increase mental health screening of their students using either version of the PHQ tool(s). Such an approach to triage may help identify those with the greatest probable need for additional mental health services. As triage needs have increased alongside service needs within college settings over the last decade (Benton et al., 2003; Gallagher, 2010), and particularly within the context of the COVID-19 pandemic (Zhai & Du, 2020), this approach may provide not only an effective method of screening, but also one which is easily and widely implemented. College counseling centers, for instance, may find particular value in the incorporation of the PHQ tool(s) into standard intake procedures given that more comprehensive psychological assessments are less frequently conducted and not as prioritized in training (Bergquist et al., 2022). Such trends may reflect broader institutional deprioritization of psychological assessment (Jenkins et al., 2022; Krishnamurthy et al., 2022). Moreover, given that the PHQ instruments assess broad internalized pathology as well as specific feelings of self-doubt, helplessness, and demoralization, our results highlight the screening utility of the PHQ instruments while also supporting the need to expand availability of more comprehensive diagnostic evaluations within college care settings for those at indicated risk.

Note

1. PHQ-9 total scores were also used to classify participants into five feigning groups using recommended interpretive guidelines of PHQ-9 total scores. A one-way ANOVA was conducted to compare groups across the MMPI-3 internalizing. Results can be seen in Supplemental Table S1.

Data Availability Statement

Data is available upon reasonable request.

Disclosure of Interest

The authors have received research support from Pearson Clinical Assessments and the University of Minnesota Press, Test Division, distributor and publisher of the Minnesota Multiphasic Personality Inventory (MMPI) family of instruments, of which the MMPI-3 belongs. No funding was obtained for this study, however. The authors report there are no other competing interests to declare.

Notes on Contributors

Nicole M. Morris' research aims to improve the diagnostic utility, implementation, and efficiency of psychological assessment, particularly amongst individuals who are transgender and gender non-conforming.

Paul B. Ingram's research focuses on advancing evidence-based assessment and barriers to engagement and retention in mental health services.

Sean M. Mitchell's research focuses on suicide risk and prevention among high-risk individuals, such as individuals with severe mental illness and individuals involved in the criminal justice system.

Sarah E. Victor's research is focused on non-suicidal self-injury (NSSI) and suicide. She is interested in identifying factors that increase risk for these problems across people, as well as intra-individual factors that contribute to short-term changes in self-harm risk in daily life.

ORCID

Nicole M. Morris http://orcid.org/0000-0001-6565-9816
Paul B. Ingram http://orcid.org/0000-0002-5409-4896
Sean M. Mitchell http://orcid.org/0000-0002-3729-7368
Sarah E. Victor http://orcid.org/0000-0002-7944-7299

References

American Psychiatric Association (APA). (1994). *Diagnostic and statistical manual of mental disorders*. (4th ed.). Author.

American Psychiatric Association (APA). (2022). *Diagnostic and statistical manual of mental disorders text revision*. (5th ed.) American Psychiatric Publishing. https://doi.org/10.1176/appi.books.9780890425787

Auerbach, R. P., Mortier, P., Bruffaerts, R., Alonso, J., Benjet, C., Cuijpers, P., Demyttenaere, K., Ebert, D. D., Green, J. G., Hasking, P., Murray, E., Nock, M. K., Pinder-Amaker, S., Sampson, N. A., Stein, D. J., Vilagut, G., Zaslavsky, A. M., & Kessler, R. C. (2018). WHO World Mental Health Surveys International College Student Project: Prevalence and distribution of mental disorders. *Journal of Abnormal Psychology*, *127*(7), 623–638. https://doi.org/10.1037/abn0000362

Ben-Porath, Y. S., & Tellegen, A. (2020a). *MMPI-3 (Minnesota Multiphasic Personality Inventory-3): Manual for administration, scoring, and interpretation*. University of Minnesota Press.

Ben-Porath, Y. S., & Tellegen, A. (2020b). *MMPI-3 (Minnesota Multiphasic Personality Inventory-3): Technical manual*. University of Minnesota Press.

Benton, S. A., Robertson, J. M., Tseng, W. C., Newton, F. B., & Benton, S. L. (2003). Changes in counseling client problems across 13 years. *Professional Psychology: Research and Practice*, *34*(1), 66–72. https://doi.org/10.1037/0735-7028.34.1.66

Bergquist, B. K., Keen, M. A., Ingram, P. B., Morris, N. M., & Schmidt, A. T. (2022). Doctoral students' intention to use assessments in their career: The incremental role of self-reported competence. *Journal of Clinical Psychology*, 1–17. https://doi.org/10.1002/jclp.23415

Boothroyd, L., Dagnan, D., & Muncer, S. (2019). PHQ-9: One factor or two? *Psychiatry Research*, *271*, 532–534. https://doi.org/10.1016/j.psychres.2018.12.048

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.) Routledge. https://doi.org/10.4324/9780203771587

Cuijpers, P., Cristea, I. A., Ebert, D. D., Koot, H. M., Auerbach, R. P., Bruffaerts, R., & Kessler, R. C. (2016). Psychological treatment of depression in college students: A metanalysis. *Depression and Anxiety*, *33*(5), 400–414. https://doi.org/10.1002/da.22461

Eisenberg, D., Gollust, S. E., Golberstein, E., & Hefner, J. L. (2007). Prevalence and correlates of depression, anxiety, and suicidality among university students. *The American Journal of Orthopsychiatry*, *77*(4), 534–542. https://doi.org/10.1037/0002-9432.77.4.534

Ferguson, C. J. (2009). An effect size primer: A guide for clinicians and researchers. *Professional Psychology: Research and Practice*, *40*(5), 532–538. https://doi.org/10.1037/a0015808

Gallagher, R. P. (2010). *National survey of counseling directors, 2010*. http://www.iacsinc.org/

Healthy Minds Network. (2019). *The healthy minds study: 2018-2019 data report*. https://healthymindsnetwork.org/wp-content/uploads/2019/09/HMS_national-2018-19.pdf

Healthy Minds Network. (2021). *The healthy minds study: 2021 Winter/Spring data report*. https://healthymindsnetwork.org/wp-content/uploads/2021/09/HMS_national_winter_2021.pdf

Ibrahim, A. K., Kelly, S. J., Adams, C. E., & Glazebrook, C. (2013). A systematic review of studies of depression prevalence in university students. *Journal of Psychiatric Research*, *47*(3), 391–400. https://doi.org/10.1016/j.jpsychires.2012.11.015

Ingram, P. B., Tarescavage, A. M., Ben-Porath, Y. S., Oehlert, M. E., & Bergquist, B. K. (2021). External correlates for the MMPI-2-Restructured Form across a national sample of veterans. *Journal of Personality Assessment*, *103*(1), 19–26. https://doi.org/10.1080/00223891.2020.1732995

Jenkins, S. R. (2022). A strong first step: Comment on Krishnamurthy et al. (2022). *Journal of Personality Assessment*, *104*(1), 23–26. https://doi.org/10.1080/00223891.2021.2001747

Kim, S., Lee, H., & Lee, K. (2020). Assessment of suicidal risk using Minnesota Multiphasic Personality Inventory-2 Restructured Form. *BMC Psychiatry*, *20*(1), 81. https://doi.org/10.1186/s12888-020-02495-2

Kotov, R., Krueger, R. F., Watson, D., Achenbach, T. M., Althoff, R. R., Bagby, R. M., Brown, T. A., Carpenter, W. T., Caspi, A., Clark, L. A., Eaton, N. R., Forbes, M. K., Forbush, K. T., Goldberg, D., Hasin, D., Hyman, S. E., Ivanova, M. Y., Lynam, D. R., Markon, K., … Zimmerman, M. (2017). The hierarchical taxonomy of psychopathology (HiTOP): A dimensional alternative to traditional nosologies. *Journal of Abnormal Psychology*, *126*(4), 454–477. https://doi.org/10.1037/abn0000258

Krishnamurthy, R., Natoli, A. P., Arbisi, P. A., Hass, G. A., & Gottfried, E. D. (2022). Professional practice guidelines for personality assessment: Response to comments by Ben-Porath (2022), Lui (2022), and Jenkins (2022). *Journal of Personality Assessment*, *104*(1), 27–29. https://doi.org/10.1080/00223891.2021.2005075

Kroenke, K., & Spitzer, R. L. (2002). The PHQ-9: A new depression diagnostic and severity measure. *Psychiatric Annals*, *32*(9), 509–515. https://doi.org/10.3928/0048-5713-20020901-06

Kroenke, K., Spitzer, R. L., & Williams, J. B. (2001). The PHQ-9: Validity of a brief depression severity measure. *Journal of General Internal Medicine*, *16*(9), 606–613. https://doi.org/10.1046/j.1525-1497.2001.016009606.x

Kroenke, K., Spitzer, R. L., & Williams, J. B. (2003). The Patient Health Questionnaire-2: Validity of a two-item depression screener. *Medical Care*, *41*(11), 1284–1292. https://doi.org/10.1097/01.MLR.0000093487.78664.3C

Levis, B., Benedetti, A., Ioannidis, J. P. A., Sun, Y., Negeri, Z., He, C., Wu, Y., Krishnan, A., Bhandari, P. M., Neupane, D., Imran, M., Rice, D. B., Riehm, K. E., Saadat, N., Azar, M., Boruff, J., Cuijpers, P., Gilbody, S., Kloda, L. A., … Thombs, B. D. (2020). Patient Health Questionnaire-9 scores do not accurately estimate depression prevalence: Individual participant data meta-analysis. *Journal of Clinical Epidemiology*, *122*, 115–128. https://doi.org/10.1016/j.jclinepi.2020.02.002

Liu, C. H., Stevens, C., Wong, S. H. M., Yasui, M., & Chen, J. A. (2019). The prevalence and predictors of mental health diagnoses and suicide among U.S. college students: Implications for addressing disparities in service use. *Depression and Anxiety*, *36*(1), 8–17. https://doi.org/10.1002/da.22830

Malpass, A., Dowrick, C., Gilbody, S., Robinson, J., Wiles, N., Duffy, L., & Lewis, G. (2016). Usefulness of PHQ-9 in primary care to determine meaningful symptoms of low mood: A qualitative study. *British Journal of General Practice*, *66*(643), e78–e84. https://doi.org/10.3399/bjgp16X683473

Manea, L., Gilbody, S., Hewitt, C., North, A., Plummer, F., Richardson, R., Thombs, B. D., Williams, B., & McMillan, D. (2016). Identifying depression with the PHQ-2: A diagnostic meta-analysis. *Journal of Affective Disorders*, *203*, 382–395. https://doi.org/10.1016/j.jad.2016.06.003

Maurer, D. M. (2012). Screening for depression. *American Family Physician*, *85*(2), 139–144. https://www.aafp.org/dam/brand/aafp/pubs/afp/issues/2012/0115/p139.pdf

McCord, D. M., & Provost, R. P. (2020). Construct validity of the PHQ-9 depression screen: Correlations with substantive scales of the MMPI-2-RF. *Journal of Clinical Psychology in Medical Settings*, *27*(1), 150–157. https://doi.org/10.1007/s10880-019-09629-z

Mitchell, A. J., Yadegarfar, M., Gill, J., & Stubbs, B. (2016). Case finding and screening clinical utility of the Patient Health Questionnaire (PHQ-9 and PHQ-2) for depression in primary care: A diagnostic metaanalysis of 40 studies. *BJPsych Open*, *2*(2), 127–138. https://doi.org/10.1192/bjpo.bp.115.001685

Moriarty, A. S., Gilbody, S., McMillan, D., & Manea, L. (2015). Screening and case finding for major depressive disorder using the Patient Health Questionnaire (PHQ-9): A meta-analysis. *General Hospital Psychiatry*, *37*(6), 567–576. https://doi.org/10.1016/j.genhosppsych.2015.06.012

Nutt, D. J. (2008). Relationship of neurotransmitters to the symptoms of major depressive disorder. *The Journal of Clinical Psychiatry*, *69*(E1), 4–7. https://www.psychiatrist.com/wp-content/uploads/2021/02/16970_relationship-neurotransmitters-symptoms-major-depressive.pdf

Santomauro, D. F., Mantilla Herrera, A. M., Shadid, J., Zheng, P., Ashbaugh, C., Pigott, D. M., Abbafati, C., Adolph, C., Amlag, J. O., Aravkin, A. Y., Bang-Jensen, B. L., Bertolacci, G. J., Bloom, S. S., Castellano, R., Castro, E., Chakrabarti, S., Chattopadhyay, J., Cogen, R. M., Collins, J. K., … Ferrari, A. J. (2021). Global prevalence and burden of depressive and anxiety disorders in 204 countries and territories in 2020 due to the COVID-19 pandemic. *The Lancet*, *398*(10312), 1700–1712. https://doi.org/10.1016/S0140-6736(21)02143-7

Spitzer, R. L., Kroenke, K., Williams, J. B. W., Patient, H. Q., & Primary Care Study, G. (1999). Validation and utility of a self-report version of PRIME-MD: The PHQ Prime Care Study. *JAMA*, *282*(18), 1737–1744. https://doi.org/10.1001/jama.282.18.1737

Tellegen, A., & Ben-Porath, Y. S. (2008). / (2011). *The Minnesota Multiphasic Personality Inventory-2 restructured form: Technical manual*. University of Minnesota Press.

Zhai, Y., & Du, X. (2020). Addressing collegiate mental health amid COVID-19 pandemic. *Psychiatry Research*, *288*, 113003. https://doi.org/10.1016/j.psychres.2020.113003