Annual Evaluation Report 2023

Prepared By:

Claudia Diaz Fuentes, Ph.D.
Economics, Team Lead

Alissa Mavridis
Evaluation Lab Fellow

Caley Marshall
Evaluation Lab Fellow

Measuring the Symptoms of Depression and Anxiety among Adolescents

Evaluation Report 2023

Executive Summary

All Faiths Children's Advocacy Center (All Faiths or AFCAC) provides services for children and families that have been affected by sexual abuse or severe physical abuse and trauma. Services include therapy, case management, community support, and wrap around services.

The Evaluation Team, comprised of the UNM Evaluation Lab and All Faiths' staff, worked between August 2022 and April 2023 to identify, adapt, and pilot a tool to measure symptoms of depression and anxiety among their adolescent clients. The Evaluation Team identified three evaluation questions;

- What instrument can All Faith's clinical staff use to track depression and anxiety symptoms among adolescents;
- What feedback and recommendations do All Faith's staff have about the screening instruments and implementation; and
- How can All Faiths use the results to support their services and processes? The Evaluation team thought it was important to include staff in this pilot and to make sure that at completion, All Faith's had a clear understanding and system for using the results.

The Evaluation Team conducted a literature review to provide a comparison of valid and reliable tools that are widely used to measure symptoms of anxiety and depression. After considering the length of the survey, ease of use, reliability and validity for adolescents ages 13-17, All Faith's leadership decided to use the Patient Health Questionnaire (PHQ-9) to measure symptoms of depression and the Generalized Anxiety Disorder (GAD-7) survey to measure symptoms of anxiety. Both of these surveys provide a diagnosis and allow providers to assess severity of symptoms. The PHQ-9 and GAD-7 were added to the Electronic Medical Record at All Faiths. Providers administered them to all clients ages 13-17 from February 23, 2023, through March 10, 2023.

After the PHQ-9 and GAD-7 were administered, the staff at All Faith's received a survey including 9 questions regarding ease of use, preference for additional training, and easy access to instrument scores. Generally, the staff felt that these surveys would be helpful in the treatment of their clients. However, additional training would make it easier for staff to administer and use the results.

Results of surveys showed a higher percentage of female clients (N=67, 83%) had moderate (31%), moderately severe (16%) and severe (15%) symptoms of depression than the male clients (N=14, 17%). Female clients also had more severe symptoms of anxiety (, whereas male clients' scores met the criteria for mild and moderate anxiety (24%). Analysis by age showed a higher percentage of clients between 15 and 16 years of age (N=36) with moderate (31%), moderately severe (17%), and severe depression (22%) compared to any other group (Ages 14 – 14, N=24, moderate 29%, moderately severe 13%, severe 8%) (Ages 17 – 18, N=21, moderate 28%, moderately severe 19%, severe 5%).

The academic literature lacks an appropriate comparison group to use as reference. The Evaluation Team identified the study by Taskina Chowdhury, and Jane Dimmit Champion called "Outcomes of Depression Screening for Adolescents Accessing Pediatric Primary Care-Based Services". (Chowdhury & Champion, 2020) This study included children ages 11-16 that visited their primary care physician, and All Faiths female clients showed more symptoms of moderate (34.78%, 18.80%) and severe (43.48%, 10.40%) depression. Males also showed more symptoms of severe depression (18.18%, 2.10%). This is not surprising due to the

client base of All Faiths. They treat children that have gone through trauma. However, it is interesting to compare them to a general population sample of the same age so they can see how they vary. Compared to a Canadian study, the study "Measurement Invariance of the GAD-7 and CESD-R-10 Among Adolescents in Canada" by Isabella Romano, of children ages 14 – 17, All Faiths clients showed higher symptoms of anxiety at all ages (Age 14, 9.39%, 6%, Age 15, 11.05%, 6.5%, Age 16, 11.93%, 6.7%, Age 17, 8.86%, 7.3%) (Romano, 2022). This study can give All Faiths a baseline when setting goals.

Although it takes time to administer the PHQ-9 and GAD-7, the Evaluation Team recommends continuing with the implementation. Based on the survey results administered to clinical staff, overall, they believe that they are helpful. The Evaluation Lab recommends additional training and guidance for the staff to use the results on a client-by-client basis. We also recommend All Faith's continues to track and review the data from these surveys to track trends and anticipate staffing and service needs.

Introduction

All Faiths Children Advocacy Center (All Faiths) "advocates for Children and Families affected by trauma." Their vision is that "New Mexico's children are safe, and all families thrive." Their focus is on the prevention, intervention, investigation, and treatment of childhood trauma. As a Children's Advocacy Center (CAC) accredited by the National Children's Alliance, they provide a child-friendly facility where law enforcement, child protection, family advocacy, therapy, and training offer children and their caregivers safety and a road to healing and recovery.

All Faiths specializes in the treatment of families, children and their caregivers struggling due to issues of divorce, homelessness, child abuse/neglect, family violence or other crises. For children, these crises are considered Adverse Childhood Experiences (ACE's). The greater number of ACEs a child experiences, the greater the probability of toxic stress and future negative health outcomes. According to the Center for Developing Child Harvard University, "fostering strong, responsive relationships between children and their caregivers, and helping children and adults build core life skills, can help to buffer a child from the effects of toxic stress. (Center on the Developing Child, Harvard University, n.d.) All Faiths services provide this buffer. See Appendix A for a map of All Faith's services.

All Faiths Clinical Director, Brittany Howell-Abbate, communicated the need for an instrument to measure and track depression and anxiety symptoms among adolescents receiving services through their Family Wellness Program (FWP). As mentioned above, All Faiths provides services for adolescents that experience trauma, and determines the level of trauma based on the number of Adverse Childhood Experiences (ACEs) clients report. Their services address these ACEs before they cultivate adverse health issues by assessing client's mental health status and symptoms of anxiety and/or depression.

The goal of this evaluation is to identify, adapt, and pilot a survey tool to measure depression and anxiety symptoms among adolescents 13 to 17 years old receiving services through the FWP at All Faiths. These services include therapy, case management, community support and wrap around services. The survey tool's goal is to measure the child's depression and anxiety level at the time of intake and throughout the recommended term of treatment whether it be therapy, case management, community support or wrap around services.

To meet this evaluation goal, the team will address these evaluations questions:

1. What instrument(s) can All Faiths' clinical staff use to track depression and anxiety symptoms among children?
2. What issues do Providers find during the pilot, and what are their recommendations to address them?
3. How can All Faiths use the results to support their services and processes?

All information regarding clients is kept in the EMR Bear (their Electronic Medical Record) in individual notes entered by the therapists. For All Faiths to collect data, they would need to manually code all individually entered notes. These notes do not have a standard form or scope. A survey tool will give them instantaneous information that can be compared internally by geological factors including a client's ACE score and services received. All Faith's average scoring results can also be compared to national average scores. This will help All Faiths staff anticipate client needs and help the organization gage the services needed by the community.

The Evaluation Team, comprised of the UNM Evaluation Lab and All Faith's leadership and staff, followed a participatory evaluation approach to identify the organization's evaluation goal. The members of the UNM Evaluation Lab participating in this evaluation are Team Lead, Claudia Diaz Fuentes, Ph.D., and Evaluation Lab Fellows Alissa Mavridis, and Cayley Marshall. Leadership from All Faiths consists of Clinical Director, Brittany Howell-Abbate, Director of CQI and Compliance, Traci Sanchez, and CEO, Krisztina Udvardi.

Work Performed

Choosing the Surveys:

The Evaluation Lab team members reviewed academic literature to identify validated scales to measure depression and anxiety symptoms. It was important to find a scale that was tested, had consistent results among children ages 13 to 17, and was available in Spanish. There are several survey or screening tools that have been validated and tested for reliability. We presented five surveys (PHQ-9, GAD-7, CES-D, RCADS, and CDI) to All Faith's and asked them to choose which tool would be most appropriate for the data they need (Martin et al., 2006; Ruby et al., 2002; Radloff, 1977; Ahlen & Ghaderi 2017). The Evaluation Lab team provided a comparison chart listing the most used scales for anxiety and depression (see Appendix B for Literature Review, Survey Summary, and Comparison Chart). After careful review, leadership at All Faiths decided that the Patient Health Questionnaire (PHQ-9) and Generalized Anxiety Disorder scale (GAD-7) were the best options. They came to this conclusion based on the number of survey questions, ease of scoring, appropriate age, and access to the survey in multiple languages. In addition, the extensive use of these two scales in many studies provides potential comparison groups in the existing literature, which can provide All Faiths with additional context. Table 1 displays the scoring and treatment actions for the PHQ-9 and Gad-7 combined into one chart. (Kroenski & Spitzer, 2002; Reetu, S. A., 2021)

Table 1. PHQ-9 and GAD 7 Scoring and Treatment Guidelines

PHQ-9 Score	GAD-7 Score	Severity	Proposed Treatment Actions
0 - 4	0 - 5	None	None
5 - 9	6 - 10	Mild	Watchful waiting, repeating at follow up.
10 - 14	11 - 15	Moderate	Consider CBT and pharmacotherapy.
15 - 19		Moderately Severe	Immediate initiation of pharmacotherapy and CBT.
20 - 27	16 - 21	Severe	Initiation of pharmacotherapy and CBT. Consider specialist referral to psychiatrist.

**Note: From "Do you Know About Major Depression Order?", S. A. Reetu, 2021, adapted from: Kroenke, K., & Spitzer, R. L. (2002). The PHQ-9: a new depression diagnostic and severity measure. *Psychiatric annals*, *32*(9), 509-515.

All Faiths added the PHQ-9 and GAD-7 to the Electronic Medical Record (EMR). A reminder alert told providers when it was necessary for the survey to be administered. During the pilot this reminder was added to all accounts, not just for clients from 13-17 years old. It was not possible to add the alert to the specific aged client records. Information regarding administration was communicated to all staff to ensure proper administration of the PHQ-9 and GAD-7.

Pilot:

The Evaluation Team presented the PHQ-9 and GAD-7 at a service providers' team meeting on December 13, 2022, to initiate a forum for open discussion. This meeting gave staff an opportunity to voice any opinions or recommendations they may have had. All Faiths staff did provide feedback; however, it was decided that surveys would be a better method of communication. Staff includes therapists, case managers, and staff managers.

All Faiths determined that the survey would be administered to all clients ages 13 to 17 receiving services through FWP from January 23, 2023 through March 10. If a child only received one service, they received the survey during that service. If the client received more than one service, the providers followed procedures as to who administered the survey once.

Using Survey Monkey, the Evaluation Lab emailed All Faiths a survey to obtain feedback on their experience administering the PHQ-9 and GAD-7 (see Appendix E for the survey questions). It included open and closed ended questions regarding the administering, scoring the survey, and whether or not they thought that it would be helpful on either an individual basis, organizational level, or both.

Data analysis:

The results show descriptive statistics for the PHQ 9, GAD-7 and the staff survey. Regarding the diagnostic surveys, results are available by severity of symptoms, score, gender, age, and in comparison to studies identified through a systematic review of the literature. Results also show a summary of main points from open ended questions from the staff survey.

Data Analysis

Sample and data collection

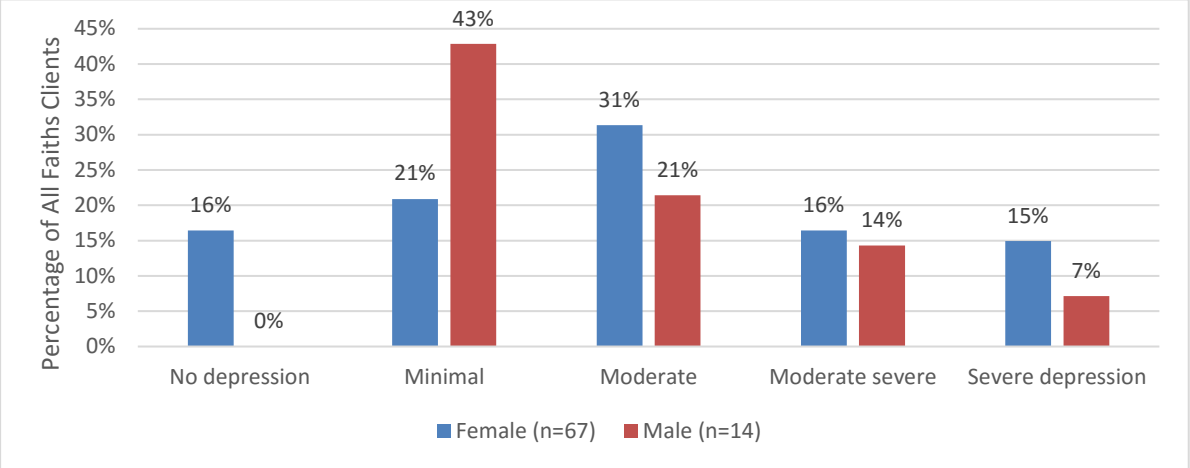
The final sample consisted of 101 All Faiths clients, 19 were outside of the target age group so they are not part of this analysis. The only exception was when we comparing All Faiths to existing literature, which included children ages 11-16. The sample final count were 81 survey responses for both the PHQ-9 and the GAD-7. All clients that received services from 02/23/23 to 03/10/23 from the age of 13-17 were given the survey.

Results:

On average, the PHQ-9 score was 12, which shows moderate symptoms of depression and it is recommended that CBT and pharmacotherapy treatment be considered. The average GAD-7 score was 10. This score shows mild symptoms of anxiety, cusping with the start of symptoms for moderate anxiety. For mild symptoms a watchful eye and repeat testing at the next appointment is recommended.

Results by Gender

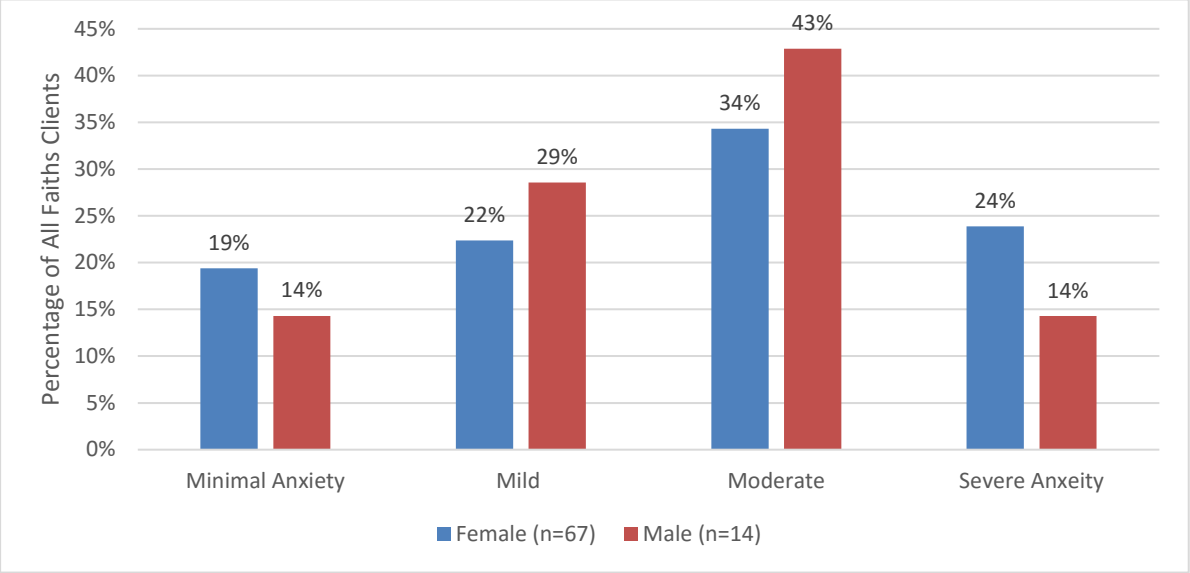
Figure 1: Percentage of clients by depression severity and gender.



Of the 81 responses, 67 were female and 14 were male (Figure 1). There are a greater number of male clients with minimal symptoms of depression (PHQ-9) than female, however female clients had more moderate, moderately severe and severe symptoms of depression.

Figure 2 shows that a larger percentage of male clients had mild (29%) and moderate (43%) anxiety compared to female clients (22% and 34%, respectively). In contrast a higher percentage of women (24%) showed severe symptoms compared to men (14%). Though the percentages by gender visually differ, statistical tests show no difference between the two groups. We expect this is due to the small sample among men and recommend oversampling to make a reliable comparison across groups.

Figure 2: Percentage of clients by anxiety severity and gender.



Results by Age

Figure 3 shows that for each age group, about 30% had moderate depression symptoms. The largest group to show moderate severe to severe depression were those 15 to 16 years of age (39%). For the 17 to 18-year-old group, nearly half of them had minimal (33%) or no depression (14%) symptoms, while moderate (29%) and moderate severe (19%) symptoms account for most of the other half. Of the 21 clients in this age group, only 1 showed severe depression (5%). Clients aged 13-14 had larger proportion of minimal (25%) or no depression symptoms (25%).

A statistical test of means shows that the only age group that statistically differs from the rest are 16-year old children (p=0.008). Among them, the average PHQ-9 score is 15.2, compared to 10.7 for the rest of the sample.

Figure 3: Percentage of clients by depression severity and age group.

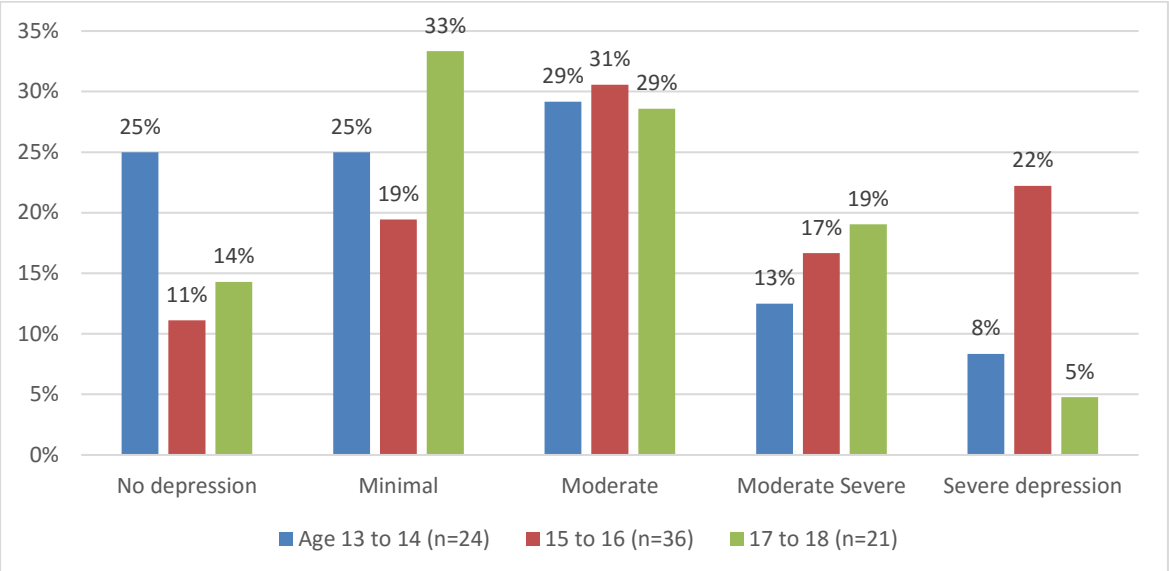
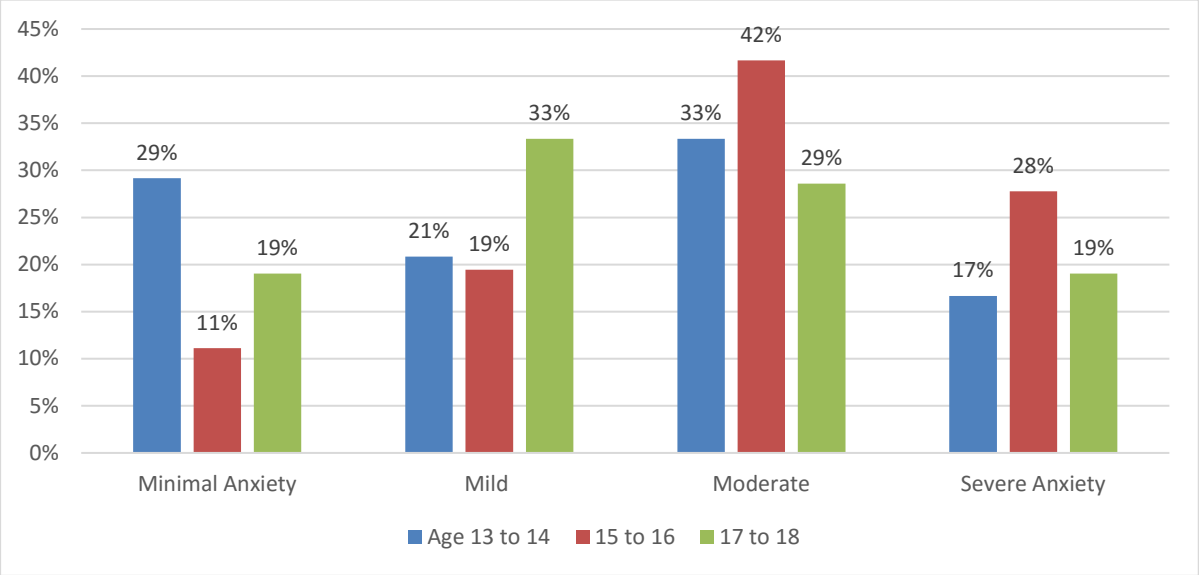


Figure 4: Number of clients by age group and anxiety severity.



Comparison to Adolescent Studies

Figure 5: Severity by ages 11 -16: Comparison to Chowdhury study. (Chowdhury & Champion, 2020)

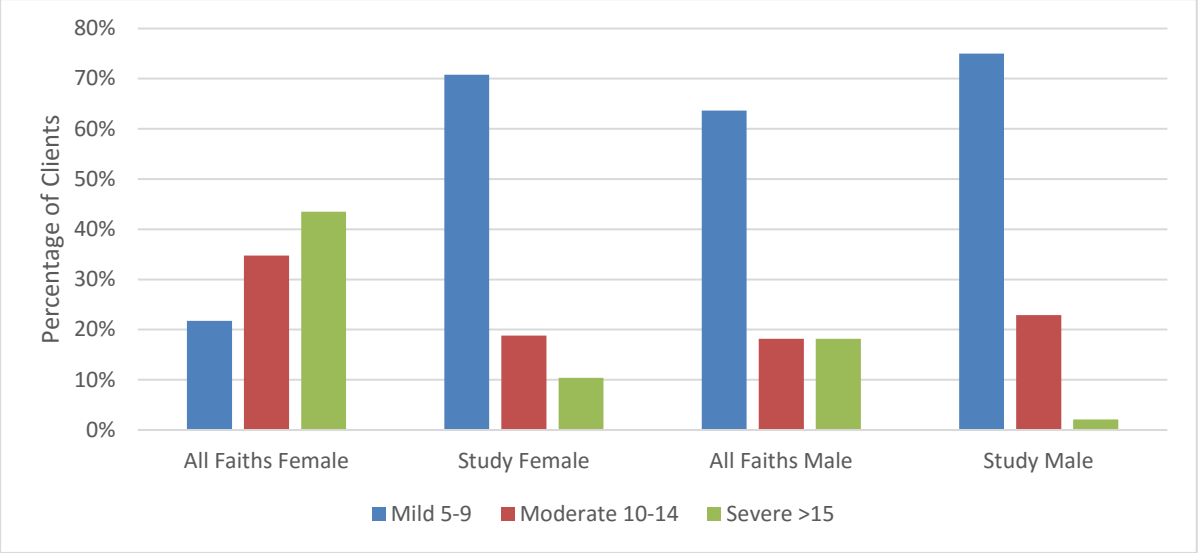


Figure 5 compares All Faith's clients scores to those used in Chowdhury & Champion (2020). The authors collected data from 11-16 years old who regularly access pediatric primary care providers in the United States. This sample represents a general population of 1,213 adolescents the same age as the All Faiths clients that completed the surveys. Comparing All Faith's results to this study, All Faiths clients a larger percentage of clients with symptoms of moderate and severe depression. This is expected since All Faith's serves clients that have experienced trauma.

Figure 6: Average GAD-7 scores by age: Comparison to Romano (2022)

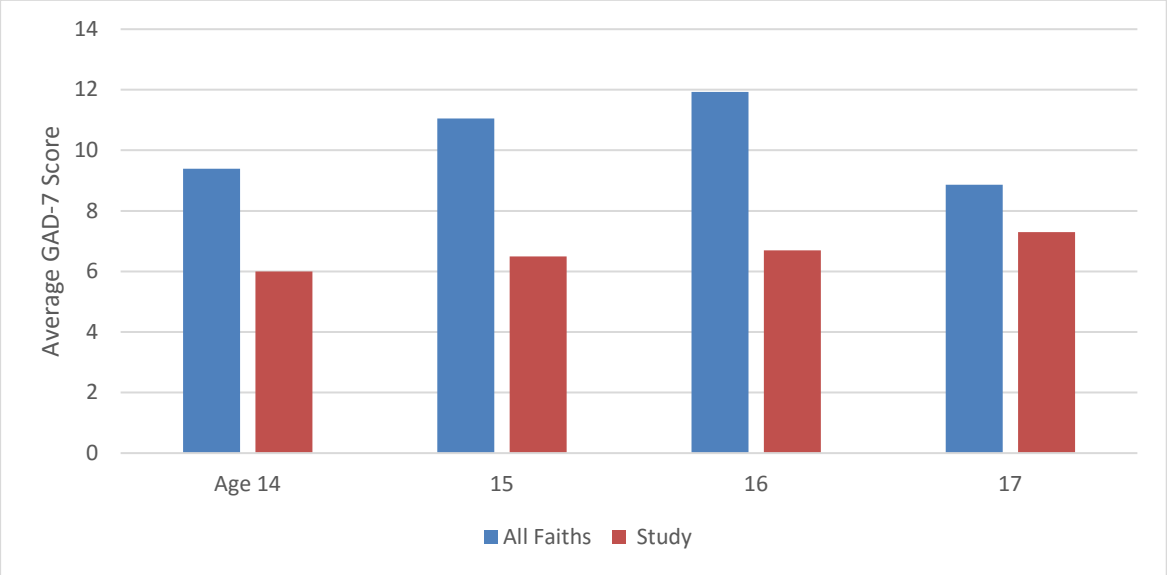


Figure 6 compares the GAD-7 scores of All Faith's clients to those from the study "Measurement Invariance of the GAD-7 and CESD-R-10 Among Adolescents in Canada" by Isabella Romano (Romano, 2022). This study used data from 59,052 adolescents from the ages of 14 to 17 attending school in Canada. All Faith's clients had more symptoms of anxiety than the same aged adolescents in this study.

Lessons from data analysis – a word of caution for future analysis:

The Evaluation Team initially expected to observe differences in scores among clients who have received services over a longer period of time compared to those with shorter periods of time. Given the severity of trauma among All Faiths clients, doing so might yield misleading results. For instance, it is possible that those receiving treatment over long periods of time do so **because** symptoms are more severe. Therefore, their scores could be higher even after a long period of treatment. The UNM Evaluation Lab team members identified this was the case once data were collected and concluded that, for a population like the one served at All Faiths, cross-sectional comparisons of this nature would not be appropriate.

Another comparison considered were ICD-10-CM codes from All Faiths' EMR. The team expected that, given the diagnosis of depression and/or anxiety, the PHQ-9 and GAD-7 could provide complementary information on specific symptoms and level of severity for each client. In other words, All Faiths could use both its ICD-10-CM codes as means to track this information, while utilizing the PHQ-9 and GAD-7 for added specificity regarding severity. This, however, was not feasible. As per the All Faith's clinical director, ICD-10-CM coding practices reflect the circumstances in which the child enters the organization. As a result, most clients (n=57, 70.3%) were assigned an F43 code "Reaction to severe stress, and adjustment disorders", followed by a depression-related code: F32 "Major Depressive Disorder, single Episode" (n=14, 17.3%) or F33 "Major Depressive Disorder, recurrent" (n=8, 9.9%). In contrast, the PHQ identified 60% of clients with moderate to severe depression. Regarding anxiety, 11 (13.6%) children received an anxiety-related code[1] and the GAD-7 identified 58% of children with moderate to severe anxiety (see Table 2).

[1] F40 "Phobic anxiety disorders", F41 "Other anxiety disorders"

Table 2. Number of children with diagnosis code.

Code	Description	Number of children with diagnosis
F43	Reaction to severe stress, and adjustment disorders	57
F32	Major depressive disorder, single episode	14
F41	Other anxiety disorders	11
F90	Attention-deficit hyperactivity disorders	9
F33	Major depressive disorder, recurrent	8
F34	Persistent mood [affective] disorders	5
Z62	Problems related to upbringing	2
Z63	Oth prob rel to prim support group, inc family circumstances	1
F95	Tic disorder	1
F91	Conduct disorders	1
F88	Other disorders of psychological development	1
F84	Pervasive developmental disorders	1
F81	Specific developmental disorders of scholastic skills	1
F80	Specific developmental disorders of speech and language	1
F71	Moderate intellectual disabilities	1
F50	Eating disorders	1
F42	Obsessive-compulsive disorder	1
F40	Phobic anxiety disorders	1
F20	Schizophrenia	1

All Faiths maintains greater specificity of client diagnosis and progress in clinical notes, which the Evaluation Lab did not access. The team concluded that ICD-10-CM diagnosis codes are not a source of information regarding the prevalence depression among teens at All Faiths and, therefore, did not consider the analysis relevant for this report.

Staff Responses to Surveys

Ten staff members responded to the feedback survey via Survey Monkey including therapists, supervisors, and case workers. The survey was conducted in order to understand their thoughts on the effectiveness of the PHQ-9 and GAD-7.

The survey asked staff members to rate how easy it was to administer the PHQ and GAD. Responses were based on a 100-point scale with 0 being not easy to administer and 100 being the easiest to administer. Respondents generally found the surveys easy to administer, with an average score of 83.9. One lower response may indicate some difficulties. The lowest score of 49 was given by a staff member who did not actually administer any of the surveys, however offered input in other open-ended questions.

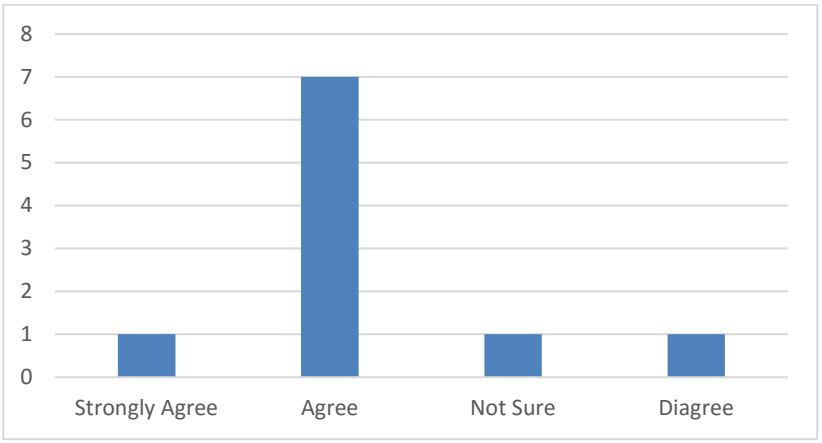
Regarding scoring, 6 out of 10 respondents estimated the scores manually. It is possible the EMR failed to provide the final score for some of the providers, or providers did not know how to find the score in the EMR since only 1 of them was able to do so. Overall, these two questions show providers would benefit from training and roleplaying practices for administering, scoring, and interpretation of results. This report provides an extensive appendix that All Faiths might find helpful for this purpose.

The survey then asked staff members to rate how easy it was for clients to understand the PHQ and GAD results. Respondents found the surveys more difficult for clients to understand than to implement. The average score for this question is 73. These results may indicate a need for additional help in this area. Again, the response to this question is on a 100-point scale with 0 being the hardest for clients to understand and 100 being the easiest.

The survey asked staff members how comfortable they felt in interpreting the results of the PHQ and GAD. Respondents were mostly comfortable in interpreting the surveys. 3 said they wanted more training. The one other response was from a case manager who had not administered any surveys. This was a close ended question with the 4 choices as listed in the graph.

The survey asked staff members if they believed that the PHQ and GAD surveys would help them with their client work. Respondents overwhelmingly believed that the surveys would be helpful with clients (n=7). Only 2 respondents disagreed that the surveys would be helpful. Their responses may be an area to explore further to better understand what is unhelpful about the surveys

Figure 7. Number of Therapists who answered "These Surveys Will Help All Faiths Better Understand Its Clients"



Finally, the survey asked staff members if they believed that the PHQ and GAD surveys would help All Faiths to better understand its clients. Respondents agreed with this statement even more. The responses to this close-ended question are displayed in Figure 7. Only one respondent disagreed, and one strongly agreed. It seems like the respondents see the surveys as more helpful for All Faiths than for themselves as therapists.

Respondents were asked two open-ended questions. The first asked them to share how the surveys might help their clients or All Faiths as a whole. Several respondents mentioned discussing or exploring symptoms using the surveys (n=3). Two respondents talked about their concerns with using the surveys, with one mentioning cultural sensitivity. One respondent said that the surveys could be used to show data trends over time for clients.

The second question asked them to share suggestions for improving the surveys. There were three responses to this question. One respondent noted concerns with how time consuming the surveys were to administer, and how they could be inappropriate for clients doing crisis management. They also noted that the scores could be misleading. One respondent simply wanted EMR bear to calculate the score itself. The last respondent noted client struggles in answering some questions, and client distress when the scores were high. They also noted that the client could see benefits from seeing changes over time.

Recommendations

These PHQ-9 and GAD-7 surveys are used to track and monitor symptoms, and can be used as a guide to help therpists, and case workers identify and treat symptoms before they worsen. According to the staff surveys, feedback from therapists was generally positive showing that most felt that the PHQ-9 and GAD-7 could be useful when treating their clients. The Evaluation Lab recommends additional training to staff, in particular to therapists, regarding scoring and use of these tools since the data they provide can be useful to track and monitor client symptoms. An additional advantage of training is understanding what these instruments do and do not do: they measure specific symptoms and provide diagnosis guidelines. They are not meant to provide a comprehensive picture of the clients' clinical needs.

Through the results we received via staff surveys, clinical staff found the PHQ-9 and GAD-7 useful, however, case managers and those providing emergent care were less inclined to think so. As mentioned above, these instruments are a useful tool, but might be better left out of emergecy cases and case management visits.

Result showed that females ages 15-16 are showing the most symptoms in depression and anxiety. This result can be used to help All Faiths focus on this population and develop treatment strategies. This trend may change over time and these surveys can help All Faiths improve their services to accommodate these changes. Currently All Faiths does not have any other unit of measurement to predict future client needs and services. Understanding what portion of clients exhibit the most symptoms can help All Faiths with decision making regarding resources.

Ongoing comparisons to general population samples can help All Faiths develop baselines for realistic goals. For instance although symptoms of anxiety for All Faiths clients aged 17 are higher than those of the Canadian study, they are closer to each other than the other ages. If 17 year olds, in general, showed a higher level of anxiety it is important for All Faiths to be aware of this and other trends. It would not be realsitic to expect results below the baseline. That said, we also recommend larger sampling for this group to obtain comparable and reliable results.

If All Faiths continues to administer these surveys it would give them an additional unit of measurement to determine how well the organization is doing over time. For this, the organization can compare results from intake throughout treatment to see improvement. Cross-sectional data like the one in this report only depicts the symptoms at that particular time, not improvement. The Evaluation Team was able to take a snap-shot of the symptoms, but these results are not a reflection of the progress of the symptoms measured throught he PHQ-9 and the GAD-7.

References

Ahlen, J., & Ghaderi, A. (2017). Evaluation of the Children's Depression Inventory—Short Version (CDI–S). *Psychological Assessment*, *29*(9), 1157.

Angold, A., Erkanli, A., Copeland, W., Goodman, R., Fisher, P. W., & Costello, E. J. (2012). Psychiatric diagnostic interviews for children and adolescents: A comparative study. Journal of the American Academy of Child & Adolescent Psychiatry, 51(5), 506-517. doi:10.1016/j.jaac.2012.02.020

Chorpita, B. F., Daleiden, E. L., Park, A. L., Ward, A. M., Levy, M. C., Cromley, T., Chiu, A. W., Letamendi, A. M., Tsai, K. H., & Krull, J. L. (2017). Child STEPs in California: A cluster randomized effectiveness trial comparing modular treatment with community implemented treatment for youth with anxiety, depression, conduct problems, or traumatic stress. *Journal of Consulting and Clinical Psychology, 85*(1), 13–25. https://doi.org/10.1037/ccp0000133

Chowdhury, T., & Champion, J. D. (2020). Outcomes of depression screening for adolescents accessing pediatric primary care-based services. *Journal of Pediatric Nursing*, *52*, 25-29.

Harvard University. (n.d.) *ACEs and Toxic Stress: Frequently Asked Questions.* https://developingchild.harvard.edu/resources/aces-and-toxic-stress-frequently-asked-questions/#graphic-text

Krause, K. R., Edbrooke-Childs, J., Singleton, R., & Wolpert, M. (2022). Are we comparing apples with oranges? Assessing improvement across symptoms, functioning, and goal progress for adolescent anxiety and depression. Child Psychiatry & Human Development, 53(4), 737-753.

Kroenke, K., & Spitzer, R. L. (2002). The PHQ-9: a new depression diagnostic and severity measure. *Psychiatric annals*, *32*(9), 509-515.

Martin, A., Rief, W., Klaiberg, A., & Braehler, E. (2006). Validity of the brief patient health questionnaire mood scale (PHQ-9) in the general population. *General hospital psychiatry*, *28*(1), 71-77.

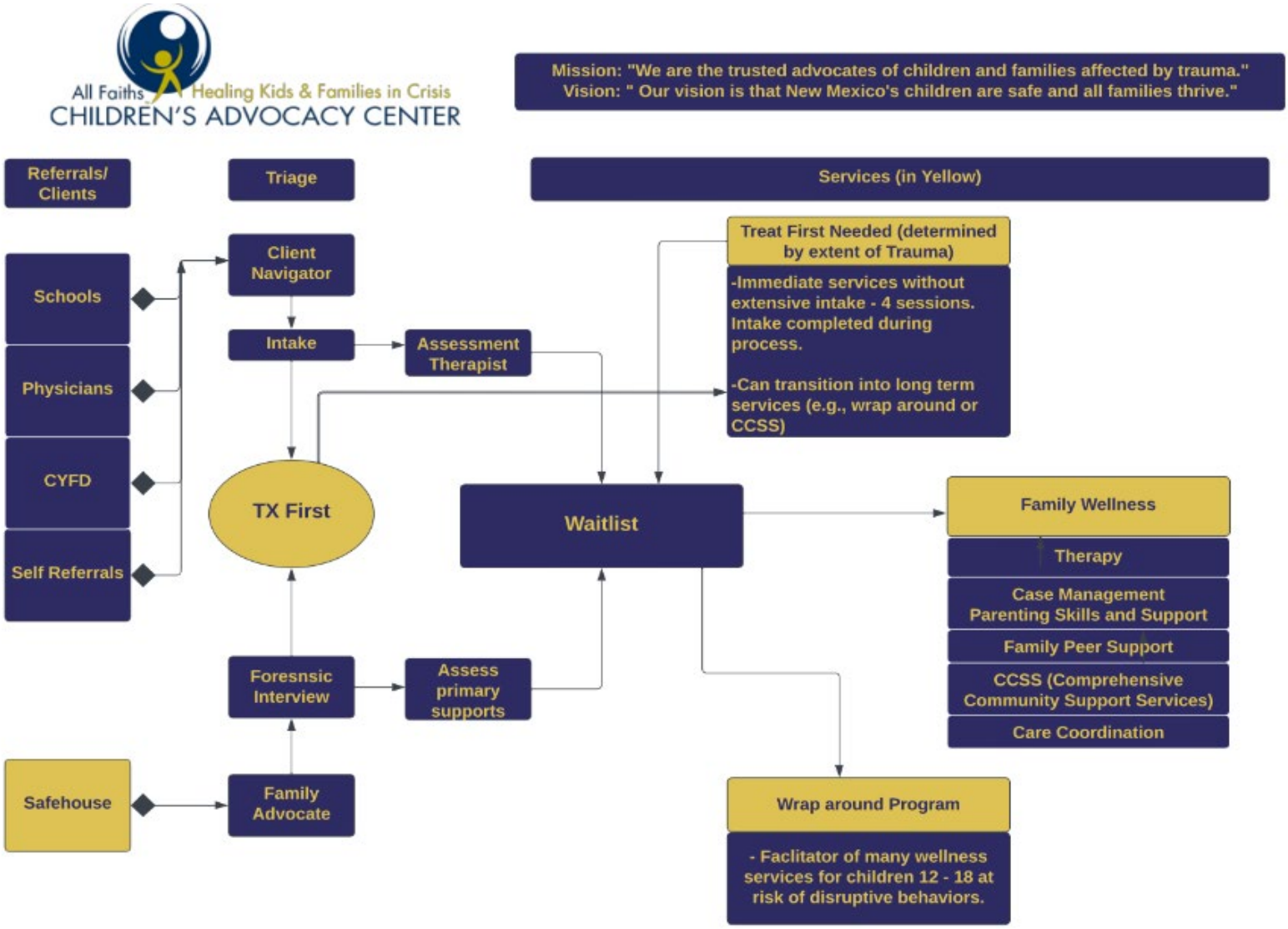
Radloff, L. S. (1977). The CES-D scale: A self-report depression scale for research in the general population. *Applied psychological measurement*, *1*(3), 385-401.

Reetu, S. A. (2021, June 21) “Do You Know About Major Depression Disorder?” Platform | CME Continuing Medical Education. https://cme.platform-med.org/2021/06/21/do-you-know-about-major-depressive-disorder.

Romano, I., Ferro, M. A., Patte, K. A., & Leatherdale, S. T. (2022). Measurement invariance of the GAD-7 and CESD-R-10 among adolescents in Canada. *Journal of pediatric psychology*, *47*(5), 585-594.

Ruby, F., da Silva, L. C., Tait, N., Rashid, A., Singleton, R., Atkins, L., Marriot, S., Dalzell, K., Labno, A., Edbrooke-Childs & Jacob, J. (2022). Children and young people’s mental health outcome measures in paediatrics. Archives of Disease in Childhood.

Appendix A – Map of All Faith's Services



AFCAC's health therapy department in the FWP consists of masters-level therapists and clinical social workers who provide evidence-based, trauma-informed behavioral health therapy. Individual, family, and group therapy are offered using the following techniques:

The goal of the UNM Evaluation Lab is to identify, adapt, and pilot a survey tool to measure depression and anxiety symptoms for children. The goal of the assessment tool is to measure the level of depression/anxiety at the time of intake and throughout the recommended term of therapy.

Evaluations Questions:

1. What instrument(s) can AFCAC clinical staff use to track depression and anxiety symptoms among children?
2. What support systems, if any, can be implemented with the survey to enable intake staff and therapists to properly administer the survey?
3. How can AFCAC use the results of symptom assessment to improve services/processes on a continued basis?

Review of the Literature

1. Ahlen, J., & Ghaderi, A. (2017). Evaluation of the Children's Depression Inventory—short version (CDI–S). Psychological Assessment, 29(9), 1157-1166. doi:10.1037/pas0000419

Summary: This study evaluates the validity and reliability of the Children's Depression Inventory short version (CDI-S) as a measure of depression and anxiety in children and adolescents. The study group consisted of children ages 8-12 in Sweden.

Research Questions:

- ✓ Does the 10 item CDI-S questionnaire have the same psychometric properties as the original 27 item CDI questionnaire?
- ✓ Does the CDI adequately measure both depression and anxiety?
- ✓ Do factors, such as gender, age and social economic status affect the result of the survey?
- ✓

Main findings:

- ✓ The CDI-S was highly correlated with the RCADS to measure depression. Although the CDI-S correlated with the SCAS scales, there was more variation when measuring anxiety.

- ✓ Girls in the study showed higher levels of internalized symptoms depression and anxiety than boys.
- ✓ Boys in the study showed higher levels of externalized symptoms of depression and anxiety than girls
- ✓ Socio-economic status was more of a factor in boys exhibiting symptoms of depression and anxiety than girls.

Methods:
The study first compared the CDI-S questionnaire to 2 widely used instruments to measure anxiety and depression that are known for their reliability and validity. The Spence Children's Anxiety Scale (SCAS) is a questionnaire used to measure anxiety symptoms. Revised Child Anxiety and Depression Scale (RCADS) is an adaption of the SCAS that was developed to better understand the interconnectedness of anxiety disorders and major depression. In this study, only the subscale of the RCADS to measure depression was used for comparison. 804 participants from schools in Sweden were given the CDI-S at the beginning of the study. The study divided the participants into two groups, those that were given the SCAS and RCADS at the same time as the CSI-S and those that were given SCAS and RCADS 2 weeks later. An internet survey was conducted for parents to report their socio-economic economic status and demographics.

Strength of Evidence:
Further research needs to be done. The study was conducted with children ages 8-12. A comparison of the RCADS data to CSI-S included data from RCADS from all children under the age of 18. The study was examined data from children that self-reported low-level symptoms of anxiety and depression. A significant portion of the parents did not report their socio-economic status.

Project Implications:
The CDI-S is less time consuming than other measures of depression and anxiety. The CDI-S measures both depression and anxiety and is more valid and reliable at measuring depression than RCADS data. Gender differences were shown to be highly significant. Socio-economic status and demographics were significant factors.

2. Angold, A., Erkanli, A., Copeland, W., Goodman, R., Fisher, P. W., & Costello, E. J. (2012). Psychiatric diagnostic interviews for children and adolescents: A comparative study. *Journal of the American Academy of Child & Adolescent Psychiatry, 51*(5), 506-517. doi:10.1016/j.jaac.2012.02.020

Summary:
This study compares 3 measures of incidences of psychiatric disorders in children and adolescents, the Diagnostic Interview Schedule for Children (DISC) ("respondent-based"), the Child and Adolescent Psychiatric Assessment (CAPA) ("interviewer-based"), and the Development and Well-Being Assessment (DAWBA) ("expert judgment"). These diagnostic interviews measure depression, anxiety, oppositional and conduct disorder,

and ADHD. The study included children ages 6-19 from Duke's Primary Care Pediatric Clinics in Durham, North Carolina, The study wishes to find if the results of measurement are similar and if not, which are the most appropriate to employ and under which circumstances.

Research Question:

- ✓ Are these 3 diagnostic interview measurements comparable and do the produce similar results?
- ✓ Does the level of training of the interviewer matter?
- ✓ Does the amount of time needed to conduct the interview matter?
- ✓ Which of the diagnostic interview measurements may over-report or under-report incidences of psychiatric disorders?

Main findings:

- ✓ DAWPA reports fewer incidences, but more severe cases. Scores were higher for DAWPA only interviews than for CAPA and DISC only interviews. DAWPA may under-report incidences. CAPA and DISC may under-report severity.
- ✓ DAWPA requires clinical training
- ✓ CAPA and DISC do not require clinical training
- ✓ CAPA only interviews and DISC only interviews had similar results.
- ✓ The DAWBA generated significantly fewer cases of depression and anxiety than the CAPA, but similar rates of behavioral disorders (ADHD, ODD, CD), and fewer cases of ADHD, ODD, and anxiety than the DISC.

Methods:

Children from 6-19 years of age with the same demographics and socio-economic level were randomly assigned into 3 groups. Each group was then given one of the three interview instruments. After one week, the participants were given another interview instrument with another interviewer.

Strength of Evidence:

This is a strong comparison. The study chose participants based on similar DMV-5 scores prior to the start of the interviews. Participant were randomly assigned, and demographics were controlled. The study did not use the subscale for phobias contained in the DISC which would report higher incidences of psychiatric disorders compared with the CAPA and the DAWBA interviews. However, given the different approaches of the interview instruments, the study was not able to control for the level of training of the interviewer, or any inherent differences in individual interviewer's biases. Due to time constraints, the study was not able to give all 3 questionnaires at the same time to participants.

Project Implications:

CAPA and DISC allow for questionnaire adjustments, such as skips, while DAWBA does not. According to the study, CAPA offers the best tracking of incidences over time. With regard to time constrains, DAWBA was completed in approximately 30 minutes. DISC required approximately 54 minutes and CAPA 60 minutes.

3. Chorpita, B. F., Daleiden, E. L., Park, A. L., Ward, A. M., Levy, M. C., Cromley, T., . . . Krull, J. L. (2017). Child steps in California: A cluster randomized effectiveness trial

comparing modular treatment with community implemented treatment for youth with anxiety, depression, conduct problems, or traumatic stress. Journal of Consulting and Clinical Psychology, 85(1), 13-25. doi:10.1037/ccp0000133

Summary: The study, conducted in Los Angeles with participants ranging in age from 5 to 15 years of age. seeks to determine if the Modular Approach to Treatment of Children (MATCH) is more effective than community-implemented treatment (CIT) in the reduction of anxiety, depression, disruptive behavior, and/or traumatic stress. MATCH-ADTC is a customized therapy approach that adjusts therapy focus based on feedback. Anxiety, depression, trauma, or conduct problems are addressed and treated. The Crisis Intervention Team (CIT) program is a widely used model for treatment. The CIT Model promotes partnerships between social services, law enforcement, behavioral health providers, courts and families and is static based on best practices for treatment. The study uses several measures to judge effectiveness:

- Brief Problem Checklist (BPC)—Child and caregiver versions
- Top problems assessment (TPA)—Child and caregiver versions
- Revised Child Anxiety and Depression Scales (RCADS)— Child and caregiver versions
- University of California at Los Angeles Post-traumatic Stress Disorder Reaction Index (UCLA PTSD Index)—Child, adolescent, and caregiver versions
- Strength and Difficulties Questionnaire (SDQ)—Child and caregiver versions
- Services assessment for children and adolescents (SACA)-- Parent version21
- Services for children and adolescents--Parent interview (SCAPI)
- Client Satisfaction Questionnaire—Child and caregiver versions

Research Question:
Which of these 2 approaches delivered better results in terms of outcomes for the deduction of disruptive behavior, and/or traumatic stress, anxiety and depression in the short-term and sustainability in the long-term?

Main findings:
Using the BPC, TPA and the other scores at baseline and at various times during invention and at final BPC and TPA scores assessment, it was found:

- ✓ 60% of participants involved in a MATCH intervention showed improvement versus 36.7% of participants involved in the CIT treatment approach. To evaluate outcomes TPA assessments were also gathered weekly.
- ✓ MATCH participants showed faster rates of improvement.
- ✓ Match participants required less follow-up therapy sessions long-term compared to CIT participants.
- ✓ Match participants had a higher rate of therapy session engagement.

Methods: 138 youth between the ages of 5 and 15 that were found to have similar cut off scores for anxiety, depression, conduct problems or traumatic stress at baseline were randomly assigned to CIT and Match approaches and randomly assigned to therapists within each cluster. Differences between the therapists in the 2 treatment groups were

not statistically significant. The final assessors of the outcomes were blinded to which treatment approach had been used.

Strength of Evidence:
This study used robust measurements to establish a baseline before intervention and used several measurements throughout the study. The weakness of the study is that MATCH does not specifically address trauma, while several of other interventions do. CIT does not specifically address anxiety. The study broke down demographics, socio-economic status and age within each treatment group.

Project Implications:
The study effectively established a baseline, used several different scoring scales and accounted for demographics and socio-economic status. Several different scoring scales may be used to evaluate the effectiveness of a program. This study also controlled for the amount of therapy sessions attended by clients. The age range in this study is consistent with the age range for evaluation in our proposed study.

4. Krause, K. R., Edbrooke-Childs, J., Singleton, R., & Wolpert, M. (2022). Are we comparing apples with oranges? Assessing improvement across symptoms, functioning, and goal progress for adolescent anxiety and depression. *Child Psychiatry & Human Development*, *53*(4), 737-753.

Summary:
This study uses 3 combinations of widely used and validated outcome scales in randomized groups. This study was conducted to see if a more meaningful and useful outcome measurement for mental health improvement and quality of life could be gleaned by using more than one indicator of improvement and by using a more holistic approach using multiple domains and measuring goal setting progress

Research Question:
Which group showed more a more meaningful a more meaningful improvement outcome? Meaningful improvement was defined as a reliable and valid improvement on a standardized scale and on an idiographic, goal-based outcome measure.

Main findings:
- ✓ Consistent cross-domain only showed meaningful improvement impact in only 15.6% of the cases. Close to one in four (24.0%) young people with reliably improved symptoms reported no reliable improvement in functioning.
- ✓ One in three (34.8%) young people reported meaningful goal progress but no reliable symptom improvement
- ✓ Symptom only measurements may over-estimate or under-estimate meaningful improvement and functionality
- ✓ Aggregate ratings may not be able to determine progress in specific and distinct symptoms indicators.

Methods:
This study analyzed naturalistic outcome date for 15,352 children aged 12-18 in England for which a diagnosis of anxiety or depression had been given after an initial assessment. The study randomly assigned the participants into 3 group to assess which group might have more meaningful and useful improvement ratings. Group 1 used two measures of internalizing symptoms (Comparison within symptom domain SDQ Emotion vs. RCADS). Group 2 used two measures of psychosocial functioning (Comparison within functioning domain SDQ Impact vs. C/ORS). Group 3 used aggregate ratings in the domains of symptoms, functioning, and assess progress towards self-defined goals (Comparison between symptoms, functioning, and goal progress domains)

Strength of Evidence:
The study was conducted longitudinally for 4 years and has a large sample size. Assignment to the groups being assessed was randomized. The study used outcome measures that are widely used and have a have been determined to be reliable and valid for determining levels of anxiety, depression, progress towards goals and externalized functionality.

Project Implications:
Child Outcome Rating Scale (CORS) which can be used with children as young a 6 years of age showed higher levels of improvement than the SDQ Impact survey. Goal progress assessments and RCADS which have many subscales for individual symptoms may be combined in a holistic approach along with CORS and the SDQ Impact survey

5. Ruby, F., da Silva, L. C., Tait, N., Rashid, A., Singleton, R., Atkins, L., ... & Jacob, J. (2022). Children and young people's mental health outcome measures in paediatrics. *Archives of Disease in Childhood*.

Summary: This paper provides a description of various mental health treatment outcome measures and provides guidance on which one to choose given the child's unique mental health diagnosis and circumstances. The paper discusses the following outcome measures:

- ✓ The Revised Children's Anxiety and Depression Scale (RCADS)
- ✓ The Patient Health Questionnaire 9 (PHQ-9)
- ✓ The Generalized Anxiety Disorder 7 survey (GAD-7)
- ✓ The Youth Self-Report (YSR)
- ✓ The Eating Disorder Examination Questionnaire (EDE-Q)

Once an outcome measure or a combination of outcome measures has been chosen, the paper provides guidance on how to interpret the findings and how the findings can inform continued mental health care.

Main findings:

- ✓ RCADS has been shown to be valid and reliable in assessing anxiety and depression. The outcome measure has 6 subscales that included separation anxiety disorder, social phobia, generalized anxiety disorder, panic disorder, obsessive compulsive disorder and low mood.

- ✓ GAD-7 is useful for measure general anxiety disorder but may not be able to capture distinct types of anxiety that the RCADS is able to measure.
- ✓ The PHQ-9 is a widely used questionnaire the is shown to be reliable in measuring depression and its severity.
- ✓ The YSR has eight subscales: the tendency to withdraw, somatic symptoms, anxiety and depression, social problems, thought problems, attention problems, rule-breaking behavior, and aggressive behavior. The subscales are grouped into externalizing and internalizing behaviors. The measure has been widely used and is available in many languages.
- ✓ The EDE-Q is not a diagnostic tool, but can give an indication of an eating disorder or chronic health problems.
- ✓ The person administering the outcome measurement should familiarize themselves with each survey and choose the one most appropriate to the child.
- ✓ Interpretations of the findings should put in the larger context surrounding the child.
- ✓ It is essential to inform the client what the measurement will be used for and to provide feedback to the client.

Project Considerations: The questionnaires do not require special training to complete. Most of the instruments discussed in this paper are used to measure outcomes of treatment for anxiety and depression were designed for youth over the age of 10. However, the RCADS is used to assess symptoms of depression and anxiety for children and young people aged 8–18 and the YSR has been used with children as young a 7. The reliability and validity of the psychometrics in the instruments was mostly evaluated on white children in the Northern Hemisphere with the exception of the YSR. More research needs to be done to test the validity for demographically diverse children and those with comorbidities, such as chronic illness.

Instrument	Age of participants	# of questions	Cost	Source	Measurement	Data Interpretation	Individual vs. Organizational Data
PHQ-9	12+	9	Free	UCLA	The PHQ-9 score ranges from 0 to 27 -Score 5–9: mild depression; -Score 10–14: moderate depression; -Score 15–19: moderately severe depression; -Score 20+: severe depression	The PHQ-9 is a widely used questionnaire that is shown to be reliable in measuring depression and its severity. (Martin, 2006)	Organizational and individual. Measures depression
GAD-7	13+	7	Free	Pfizer	The GAD-7 score ranges from 0 to 21. -Score 0-4: Minimal Anxiety. -Score 5-9: Mild Anxiety. -Score 10-14: Moderate Anxiety. -Score 15+: Severe Anxiety.	GAD-7 is useful for measure general anxiety disorder but may not be able to capture distinct types of anxiety that the RCADS is able to measure. (Ruby, 2022)	Organizational and individual Measures anxiety
CES-D	6+	20	Free	Laurie Radloff	The CES-D score ranges from 0 to 60. -If more than four questions are missing ansv do not score the CES-D questionnaire. -A score of 16 points or more is considered depressed.	The CES-D scale is designed to measure depressive symptomatology in the general population. (Radloff, 1977)	Individual Measures Depression
CDI	7 - 17	28	Pricing Varies	Pearson Assessment	The CDI score ranges from 0 to 54. -A higher CDI score means a higher depressive state.	The CDI-S is less time consuming than other measures of depression and anxiety. The CDI-S measures both depression and anxiety and is more valid and reliable at measuring depression than RCADS data. (Ahlen, 2017)	Individual Measures anxiety and depression
RCAD S	6 - 18	48	Free	Chorpita and Colleagues	To score the RCADS manually, each item is assigned a numerical value from 0-3, where 0 = Never, 1 = Sometimes, 2 = Often, and 3 = Always.	RCAD S	6 - 18

AFCAC Feedback Survey **The UNM Evaluation Lab is supporting All Faiths' pilot of the PHQ-9 and GAD-7 as instruments to track symptoms of depression and anxiety. The pilot data collection stage is complete! that means we are ready to learn how it went, so we really appreciate your feedback. Only the Evaluation Lab staff will see your answers, which will remain anonymous. We will analyze the data and share a summary report. Please feel free to reach out to Claudia Diaz at claudiadf@unm.edu if you have any questions!**

1. How many of your clients completed the depression and anxiety surveys?

◯ None (scroll down and select Done!) ◯ 1 or more

◯ I don't remember if I did this

2. How easy was for you to administer the the depression and anxiety symptoms surveys? (choose from 0 to 100)

0 - Extremely hard 100 - Extremely easy

3. How easy was it for clients to understand the survey?

0 - Extremely difficult 100 - Extremely easy

4. Regarding the scores of the depression and anxiety surveys, please choose the option that applies to you:

◯ The EMR provided a final score for each client ◯ I estimated the final score for each client

◯ I don't remember/don't know

5. Choose the option that best applies to you

☐ I am comfortable interpreting the results.

☐ I have an idea of what the results mean, but more training would be helpful. I'm not comfortable interpreting the results.

☐ Other (please specify)

6. How much would you agree with the following statements?

	Strongly agree	Agree	Disagree	Strongly Disagree	I'm not sure
These surveys will help me with my client work.	○	○	○	○	○
These surveys will help All Faiths better understand its clients.	○	○	○	○	○

7. If you think it would help, please share how the surveys could help your client and/or All Faiths?

8. What is your role at AFCAC? Please select the option that closest applies to you.

- ○ Case Manager
- ○ Therapist
- ○ Wrap Around
- ○ CSW
- ○ Supervisor

9. Please share any suggestions to improve the depression and anxiety surveys.

[Page left blank on purpose]

Patient Health Questionnaire and General Anxiety Disorder (PHQ-9 and GAD-7)

Date_____ Patient Name:_____ Date of Birth:

Over the last 2 weeks, how often have you been bothered by any of the following problems? Please circle your answers.

PHQ-9	Not at all	Several days	More than half the days	Nearly every day
1. Little interest or pleasure in doing things.	0	1	2	3
2. Feeling down, depressed, or hopeless.	0	1	2	3
3. Trouble falling or staying asleep, or sleeping too much.	0	1	2	3
4. Feeling tired or having little energy.	0	1	2	3
5. Poor appetite or overeating.	0	1	2	3
6. Feeling bad about yourself – or that you are a failure or have let yourself or your family down.	0	1	2	3
7. Trouble concentrating on things, such as reading the newspaper or watching television.	0	1	2	3
8. Moving or speaking so slowly that other people could have noticed. Or the opposite – being so fidgety or restless that you have been moving around a lot more than usual.	0	1	2	3
9. Thoughts that you would be better off dead, or of hurting yourself in some way.	0	1	2	3
Add the score for each column				

Total Score (add your column scores):

If you checked off any problems, how difficult have these made it for you to do your work, take care of things at home, or get along with other people? (Circle one)

Not difficult at all **Somewhat difficult** **Very Difficult** **Extremely Difficult**

Over the last 2 weeks, how often have you been bothered by any of the following problems? Please circle your answers.

GAD-7	Not at all sure	Several days	Over half the days	Nearly every day
1. Feeling nervous, anxious, or on edge.	0	1	2	3
2. Not being able to stop or control worrying.	0	1	2	3
3. Worrying too much about different things.	0	1	2	3
4. Trouble relaxing.	0	1	2	3
5. Being so restless that it's hard to sit still.	0	1	2	3
6. Becoming easily annoyed or irritable.	0	1	2	3
7. Feeling afraid as if something awful might happen.	0	1	2	3
Add the score for each column				

Total Score (add your column scores):

If you checked off any problems, how difficult have these made it for you to do your work, take care of things at home, or get along with other people? (Circle one)

Not difficult at all **Somewhat difficult** **Very Difficult** **Extremely Difficult**

UHS Rev 4/2020

INSTRUCTION MANUAL
Instructions for Patient Health Questionnaire (PHQ) and GAD-7 Measures

BACKGROUND

The Primary Care Evaluation of Mental Disorders (PRIME-MD) was an instrument developed and validated in the early 1990s to efficiently diagnose five of the most common types of mental disorders presenting in medical populations: depressive, anxiety, somatoform, alcohol, and eating disorders.[1] Patients first completed a one-page 27-item screener and, for those disorders for which they screened positive, were asked additional questions by the clinician using a structured interview guide. However, this 2-stage process took an average of 5-6 minutes of clinician time in patients without a mental disorder diagnosis and 11-12 minutes in patients with a diagnosis. This proved to be a barrier to use given the competing demands in busy clinical practice settings.

Therefore, in two large studies enrolling 6000 patients (3000 from general internal medicine and family practice clinics and 3000 from obstetrics-gynecology clinics), a self-administered version of the PRIME-MD called the Patient Health Questionnaire (PHQ) was developed and validated.[2,3] In the past decade, the PHQ in general and the PHQ-9 depression scale in particular [4-6] have gained increasing use in both research and practice. The original PRIME-MD is now largely of historical interest and seldom used except in a few types of research studies.

Given the popularity of the PHQ-9 for assessing and monitoring depression severity, a new 7-item anxiety scale using a response set similar to the PHQ-9 was initially developed to diagnose generalized anxiety disorder (hence its name, the GAD-7) and validated in 2740 primary care patients.[7] Though originally developed to diagnose generalized anxiety disorder, the GAD-7 also proved to have good sensitivity and specificity as a screener for panic, social anxiety, and post- traumatic stress disorder.[8] Finally, the PHQ-15 was derived from the original PHQ studies and is increasingly used to assess somatic symptom severity and the potential presence of somatization and somatoform disorders.[9]

Each PHQ module can be used alone (e.g. the PHQ-9 if depression is the condition of interest), together with other modules, or as part of the full PHQ. Also, alternative or abbreviated versions of the PHQ-9 and GAD-7 are sometimes used in certain screening or research settings [10-14] Although the PHQ was originally developed to detect five disorders, the depression, anxiety, and somatoform modules (in that order) have turned out to be the most popular.[10] Also, most primary care patients with depressive or anxiety disorders present with somatic complaints and co- occurrence of somatic, anxiety, and depressive symptoms (the *SAD* triad) is exceptionally common. This is the rationale behind the PHQ-SADS screener.[15] The most commonly used versions of the PHQ scales are summarized in **Table 1, page 3**.

CODING AND SCORING

The full PHQ, Brief PHQ, and PHQ for Adolescents (PHQ-A) can be used to establish provisional diagnoses for selected DSM-IV disorders. The diagnostic algorithm for the PHQ modules are included in footers at the bottom of each page of the PHQ, and also reiterated in **Table 2, page 4**. The other measures are principally used to derive severity scores (PHQ-9 and PHQ-8 for depressive symptom severity; GAD-7 for anxiety symptom severity; PHQ-15 for somatic symptom severity) or as ultra-brief screeners (PHQ-2, GAD-2, PHQ-4). An example in which the PHQ depression module can be used as both a diagnostic module as well as a depression severity score (PHQ-9 score) is shown in **Table 3, page 5.**

Over time, the severity scores have been a particularly popular use of the measures, and are now used much more commonly than the provisional diagnoses. For example, cutpoints of 5, 10, and 15 represent mild, moderate, and severe levels of depressive, anxiety, and somatic symptoms, on the PHQ-9, GAD-7, and PHQ-15 respectively. Also, a cutpoint of 10 or greater is considered a ―yellow flag‖ on all 3 measures (i.e., drawing attention to a possible clinically significant condition), while a cutpoint of 15 is a ―red flag‖ on all 3 measures (i.e., targeting individuals in whom active treatment is probably warranted). For the ultra-brief measures (PHQ-2 and GAD-2), a score of 3 or greater should prompt administration of the full PHQ-9 and/or GAD-7, as well as a clinical interview to determine whether a mental disorder is present.

The final question on the PHQ (and some of its abbreviated versions) asks the patients to report ―how difficult have these problems made it for you to do your work, take care of things at home, or get along with other people?‖ This single patient-rated difficulty item is not used in calculating any PHQ score or diagnosis but rather represents the patient's global impression of symptom-related impairment. It may be useful in decisions regarding initiation of or adjustments to treatment since it is strongly associated with both psychiatric symptom severity as well as multiple measures of impairment and health-related quality of life.

A particularly important question is how to assess suicide risk in individuals who answer positively to the 9th question of the PHQ-9. A four-item screener has been developed that may assist in positive responses to this 9th question [16], although a final decision about the actual risk of self-harm requires a clinical interview.

Table 1. Versions: Patient Health Questionnaire (PHQ) Family of Measures

Measure	Description	Scoring	References
Core			
PRIME-MD	Predecessor of PHQ, now mainly of historical interest.	Combined self-administered patient screener with clinician follow-up questions.	1
PHQ	Five modules covering 5 common types of mental disorders: depression, anxiety, somatoform, alcohol, and eating.	Selected (but provisional) DSM-IV diagnoses for all types of disorders except somatoform.	2, 3
PHQ-9	Depression scale from PHQ.	Nine items, each of which is scored 0 to 3, providing a 0 to 27 severity score.	1, 4, 5, 6, 10
GAD-7	Anxiety measure developed after PHQ but incorporated into PHQ-SADS.	Seven items, each of which is scored 0 to 3, providing a 0 to 21 severity score.	7, 8, 10
PHQ-15	Somatic symptom scale from PHQ.	Fifteen items, each of which is scored 0 to 2, providing a 0 to 30 severity score.	9, 10
PHQ-SADS	PHQ-9, GAD-7, and PHQ-15 measures, plus panic measure from original PHQ.	See scoring for these scales above.	10
Variants			
Brief PHQ	PHQ-9 and panic measures from original PHQ plus items on stressors and women's health.	See scoring for PHQ above. Stressor and women's health items are not diagnostic or scored.	3
PHQ-A	Substantially modified version of PHQ developed for use in adolescents. Moderate data exists for validity but much less than for original PHQ.	Diagnostic scoring described in manual, available upon request.	11
PHQ-2	First 2 items of PHQ-9. Ultra-brief depression screener.	Two items scored 0 to 3 (total score of 0-6)	10, 12
GAD-2	First 2 items of GAD-7. Ultra-brief anxiety screener.	Two items scored 0 to 3 (total score of 0-6)	8, 10, 12
PHQ-4	PHQ-2 and GAD-2.	See PHQ-2 and GAD-2 above.	10, 12, 13
PHQ-8	All items of PHQ-9 except the 9th item on self-harm. Mainly used in non-depression research studies.	Eight items, each of which is scored 0 to 3, providing a 0 to 24 severity score.	5, 10, 14

Table 2. Diagnostic Algorithms for the PHQ

Page 1

Somatoform Disorder if at least 3 of #1a-m bother the patient –a lot‖ and lack an adequate biological explanation.

Major Depressive Syndrome if #2a or b and five or more of #2a-i are at least —More than half the days‖ (count #2i if present at all) .

Other Depressive Syndrome if #2a or b and two, three, or four of #2a-i are at least —More than half the days‖ (count #2i if present at all).

Note**:** the diagnoses of Major Depressive Disorder and Other Depressive Disorder requires ruling out normal *bereavement (mild symptoms, duration less than 2 months***)**, a history of a *manic* episode (Bipolar Disorder) and a *physical disorder, medication or other drug* as the biological cause of the depressive symptoms.

Page 2

Panic Syndrome if #3a-d are all ‗YES' and 4 or more of #4a-k are ‗YES'.

Other Anxiety Syndrome if #5a and answers to three or more of #5b-g are —More than half the days‖.

Note**:** The diagnoses of Panic Disorder and Other Anxiety Disorder require ruling out a *physical disorder***,** *medication or other drug* as the biological cause of the anxiety symptoms.

Page 3

Bulimia Nervosa if #6a,b, and c and #8 are ‗YES';

Binge Eating Disorder the same but #8 is either ‗NO' or left blank.

Alcohol abuse if any of #10a-e are –YES‖.

Additional Clinical Considerations. After making a provisional diagnosis with the PHQ, there are additional clinical considerations that may affect decisions about management and treatment.

- *Have current symptoms been triggered by psychosocial* ***stressor(s****)?*
- *What is the* ***duration*** *of the current disturbance and has the patient received any* ***treatment*** *for it?* • *To what extent are the patient's symptoms i****mpairing*** *his or her usual work and activities?*
- *Is there a* ***history*** *of similar episodes, and were they* ***treated****?*
- *Is there a* ***family history*** *of similar conditions?*

Table 3. Example of PHQ Depression Module for both Diagnostic and Severity Purposes

Patient: A 43-year-old woman who looks sad and complains of fatigue for the past month.

2. Over the last 2 weeks, how often have you been bothered by any of the following:	**Not at all**	**Several days**	**More than half the days**	**Nearly every day**
	(0)	(1)	(2)	(3)
a. Little interest or pleasure in doing things?.....	☐	☐	☐	☒
b. Feeling down, depressed, or hopeless?.....	☐	☒	☐	☐
c. Trouble falling or staying asleep, or sleeping too much?	☐	☐	☒	☐
d. Feeling tired or having little energy?.....	☐	☐	☐	☒
e. Poor appetite or overeating?.....	☐	☒	☐	☐
f. Feeling bad about yourself—or that you are a failure or have let yourself or your family down?.....	☐	☐	☒	☐
g. Trouble concentrating on things, such as reading the newspaper or watching television?	☐	☐	☐	☒
h. Moving or speaking so slowly that other people could have noticed? Or the opposite—being so fidgety or restless that you have been moving around a lot more than usual?.....	☒	☐	☐	☐
i. Thoughts that you would be better off dead or of hurting yourself in some way?	☐	☒	☐	☐

FOR OFFICE CODING: Maj Dep Syn if #2a or b and five or more of #2a-i are at least —More than half the days‖ (count #2i if present at all) . Other Dep Syn if #2a or b and two, three, or four of #2a-i are at least —More than half the days‖ (count #2i if present at all).

Major Depressive Disorder Diagnosis. The criteria for Major Depressive Syndrome are met since she checked #2a ‑nearly every day‖ and five of items #2a to i were checked ‑more than half the days‖ or ‑nearly every day‖. Note that #2i, suicidal ideation, is counted whenever it is present.

In this case, the diagnosis of Major Depressive Disorder (not Syndrome) was made since questioning by the physician indicated no history of a manic episode; no evidence that a physical disorder, medication, or other drug caused the depression; and no indication that the depressive symptoms were normal bereavement. Questioning about the suicidal ideation indicated no significant suicidal potential.

PHQ-9 Depression Severity. This is calculated by assigning scores of 0, 1, 2, and 3, to the response categories of ‑not at all,‖ ‑several days,‖ ‑more than half the days,‖ and ‑nearly every day,‖ respectively. PHQ-9 total score for the nine items ranges from 0 to 27. In the above case, the PHQ-9 depression severity score is 16 (3 items scored 1, 2 items scored 2, and 3 items scored 3).

Scores of 5, 10, 15, and 20 represent cutpoints for mild, moderate, moderately severe and severe depression, respectively. Sensitivity to change has also been confirmed.

USE OF SOME SCREENERS AS SEVERITY AND OUTCOME MEASURES

PHQ-9 Depression Severity. This is calculated by assigning scores of 0, 1, 2, and 3, to the response categories of ‑not at all,‖ ‑several days,‖ ‑more than half the days,‖ and ‑nearly every day,‖ respectively. PHQ-9 total score for the nine items ranges from 0 to 27. In the above case (see table 3, page 5), the PHQ-9 depression severity score is 16 (3 items scored 1, 2 items
scored 2, and 3 items scored 3). Scores of 5, 10, 15, and 20 represent cutpoints for mild, moderate, moderately severe and severe depression, respectively. Sensitivity to change has also been confirmed. The **PHQ-8** is scored just like the PHQ-9 and its total score ranges from 0 to 24. Cutpoints on the PHQ-8 are identical to the PHQ-9.

GAD-7 Anxiety Severity. This is calculated by assigning scores of 0, 1, 2, and 3, to the response categories of ‑not at all,‖ ‑several days,‖ ‑more than half the days,‖ and ‑nearly every day,‖ respectively. GAD-7 total score for the seven items ranges from 0 to 21. Scores of 5, 10, and 15 represent cutpoints for mild, moderate, and severe anxiety, respectively. Though designed primarily as a screening and severity measure for generalized anxiety disorder, the GAD-7 also has moderately good operating characteristics for three other common anxiety disorders – panic disorder, social anxiety disorder, and post-traumatic stress disorder. When screening for anxiety disorders, a recommended cutpoint for further evaluation is a score of 10 or greater.

PHQ-2 and GAD-2 Severity. These consist of the first two items of the PHQ-9 and GAD-7 respectively, and constitute the two core DSM-IV items for major depressive disorder and generalized anxiety disorder, respectively. Each ranges from a score of 0 to 6. The operating characteristics of these ultra-brief measures are quite good; the recommended cutpoints for each when used as screeners is a score of 3 or greater. When used together, they are referred to as the **PHQ-4** a 4-item screening measure which ranges from a score of 0 to 12, and serves as a good measure of ‑caseness‖ (i.e., the higher the score, the more likely there is an underlying depressive or anxiety disorder). In particular, the PHQ-2 and GAD-2 subscores of the PHQ-4 provide separate depressive and anxiety scores, and can be used as screeners for depression and anxiety.

PHQ-15 Somatic Symptom Severity. This is calculated by assigning scores of 0, 1, and 2 to the response categories of ‑not at all‖, ‑bothered a little‖, and ‑bothered a lot‖, for the 13 somatic symptoms of the PHQ (items 1a-1m). Also, 2 items from the depression module (sleep and tired) are scored 0 (‑not at all‖), 1 (‑several days‖) or 2 (‑more than half the days‖ or ‑nearly every day‖). Thus, a PHQ-15 score can be derived from page 1 of the PHQ, or from separate administration of the PHQ-15 scale or the PHQ-SADS. PHQ-15 scores of 5, 10, and 15 represent cutpoints for low, medium, and high somatic symptom severity, respectively.

Sensitivity to Change for Monitoring Treatment Outcomes. A particularly important use of a measure is its responsiveness to changes of condition severity over time. This is well- established for the PHQ-9 which is increasingly used as a measure to assess the level of depression severity (for initial treatment decisions) as well as an outcome tool (to determine treatment response).[6,10] An example of how different PHQ-9 severity levels might guide treatment is shown in **Table 4, page 7**. There is preliminary evidence that the PHQ-15 may be responsive to changes as individuals with somatoform disorders or high somatization are treated.[10] The GAD-7 has demonstrated change as a secondary anxiety outcome in several depression trials, but has not yet been studied as a primary outcome in anxiety trials. Also, since there is more diagnostic splitting for anxiety than for depressive disorders, it remains to be determined whether a single anxiety measure can suffice as an outcome measure. It is likely the GAD-7 will be useful but not yet certain it will be sufficient.

Psychometrics. The psychometrics of the PHQ and its component scales are described in the validation articles for specific measures (see Selected References on page 9) and are summarized in a review article on the PHQ-9, GAD-7, and PHQ-15.[10]

Table 4. PHQ-9 Scores and Proposed Treatment Actions *

PHQ-9 Score	Depression Severity	Proposed Treatment Actions
0 – 4	None-minimal	None
5 – 9	Mild	Watchful waiting; repeat PHQ-9 at follow-up
10 – 14	Moderate	Treatment plan, considering counseling, follow-up and/or pharmacotherapy
15 – 19	Moderately Severe	Active treatment with pharmacotherapy and/or psychotherapy
20 – 27	Severe	Immediate initiation of pharmacotherapy and, if severe impairment or poor response to therapy, expedited referral to a mental health specialist for psychotherapy and/or collaborative management

* From Kroenke K, Spitzer RL, Psychiatric Annals 2002;32:509-521

TRANSLATIONS

There are numerous translations of the PHQ as well as the PHQ-9 and GAD-7 available in many languages, which are freely downloadable on the PHQ website (www.phqscreeners.com). The abbreviated versions of these measures – PHQ-8, PHQ-2, GAD-2, and PHQ-4 – can simply be derived from the translations by selecting the relevant items (see Table 1, page 3). The PHQ-15 can also be simply derived by selecting the 13 somatic items (1a-1m), plus the *sleep* and *tired* items (2c and 2c) from the PHQ translations.

Many of the translations have been developed by the MAPI Research Institute using an internationally accepted translation methodology. Thus, most of the translations are linguistically valid. However, unlike the English versions of the PHQ and GAD-7, few of the translations have been psychometrically validated against an independent structured psychiatric interview.

WEBSITE

Copies of the PHQ family of measures, including the GAD-7, are available at the website:

www.phqscreeners.com

Also, translations, a bibliography, an instruction manual, and other information is provided on this website.

QUESTIONS REGARDING DEVELOPMENT, ACKNOWLEDGMENTS AND USE

The PHQ family of measures (see Table 1, page 3), including abbreviated and alternative versions as well as the GAD-7, were developed by Drs. Robert L. Spitzer, Janet B.W. Williams, Kurt Kroenke and colleagues, with an educational grant from Pfizer Inc.

All of the measures included in Table 1 are in the public domain. No permission is required to reproduce, translate, display or distribute.

SELECTED REFERENCES

1. Spitzer RL, Williams JBW, Kroenke K, Linzer M, deGruy FV, Hahn SR, Brody D, Johnson JG. Utility of a new procedure for diagnosing mental disorders in primary care: The PRIME-MD 1000 study. JAMA 1994;272:1749-1756.
2. Spitzer RL, Kroenke K, Williams JBW, for the Patient Health Questionnaire Primary Care Study Group. Validation and utility of a self-report version of PRIME-MD: the PHQ Primary Care Study. JAMA 1999;282:1737-1744.
3. Spitzer RL, Williams JBW, Kroenke K, et al. Validity and utility of the Patient Health Questionnaire in assessment of 3000 obstetrics-gynecologic patients. Am J Obstet Gynecol 2000; 183:759-769
4. Kroenke K, Spitzer RL, Williams JBW. The PHQ-9: Validity of a brief depression severity measure. J Gen Intern Med 2001;16:606-613.
5. Kroenke K, Spitzer RL. The PHQ-9: a new depression diagnostic and severity measure. Psychiatric Annals 2002;32:509-521. *[also includes validation data on PHQ-8*]
6. Löwe B, Unutzer J, Callahan CM, Perkins AJ, Kroenke K. Monitoring depression treatment outcomes with the Patient Health Questionnaire-9. Med Care 2004;42:1194-1201
7. Spitzer RL, Kroenke K, Williams JBW, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med 2006;166:1092-1097.
8. Kroenke K, Spitzer RL, Williams JBW, Monahan PO, Löwe B. Anxiety disorders in primary care: prevalence, impairment, comorbidity, and detection. Ann Intern Med 2007;146:317-325. *[validation data on GAD-7 and GAD-2 in detecting 4 common anxiety disorders)]*
9. Kroenke K, Spitzer RL, Williams JBW. The PHQ-15: Validity of a new measure for evaluating somatic symptom severity. Psychosom Med 2002;64:258-266.
10. Kroenke K, Spitzer RL, Williams JBW, Löwe B. The Patient Health Questionnaire somatic, anxiety, and depressive symptom scales: a systematic review. Gen Hosp Psychiatry 2010 (in press).
11. Johnson JG, Harris ES, Spitzer RL, Williams JBW. The Patient Health Questionnaire for Adolescents: Validation of an instrument for the assessment of mental disorders among adolescent primary care patients. J Adolescent Health. 2002;30:196-204.
12. Kroenke K, Spitzer RL, Williams JBW. The Patient Health Questionnaire-2: validity of a two-item depression screener. Med Care 2003; 41:1284-1292.
13. Kroenke K, Spitzer RL, Williams JBW, Löwe B. An ultra-brief screening scale for anxiety and depression: the PHQ-4. Psychosomatics 2009;50:613-621
14. Kroenke K, Strine TW, Spitzer RL, Williams JBW, Berry JT, Mokdad AH. The PHQ-8 as a measure of current depression in the general population. J Affective Disorders 2009;114:163-173.

15. Löwe B, Spitzer RL, Williams JBW, Mussell M, Schellberg D, Kroenke K. Depression, anxiety, and somatization in primary care: syndrome overlap and functional impairment. Gen Hosp Psychiatry 2008;30:191-199.

16. Dube P, Kroenke K, Bair MJ, Theobald D, Williams L. The P4 screener: a brief measure for assessing potential suicidal risk. J Clin Psychiatry Primary Care Companion 2010 (in press). *[Algorithm for following up on positive responses to 9th item of PHQ-9]*