Literacy limitations to psychological evaluation tools: The case of MU*

Cynthia Luo[1], Claire Ufongene[1] and Jennifer Weintraub[2]

Introduction to the reader

In this case, a survivor of torture presents with symptoms clinically consistent with both major depressive disorder (MDD) and post-traumatic stress disorder (PTSD). During her evaluation, a validated psychological questionnaire for PTSD was administered verbally through a translator and accurately identified this diagnosis. However, a self-administered (read and completed by the client) questionnaire for MDD vastly underestimated the severity of her symptoms and failed to diagnose her with depression. The client had not completed grade school, so it is likely that her literacy level impacted the accuracy of this questionnaire. This highlights one of the many limitations that exist when administering psychological surveys. Through understanding these limitations, forensic evaluators can develop ways to identify, mitigate, and overcome limitations of these useful tools.

Background

Patient MU is an approximately 35-year-old female seeking asylum in the United States after experiencing years of gender-based violence in Central America. Growing up, she attended school through the third grade before her father removed her from further education because he believed education was not for girls. As a young child, MU was sexually abused by her father's co-workers, a relative on her mother's side, and later her own father.

In early adulthood, MU endured ongoing psychological, physical, and sexual abuse from her long-time partner. This included being hit in the head to the point of losing consciousness and being forcibly confined to the house for one month. MU reported this abuse to the local police multiple times, but they never took her partner into custody and told her they could not help her. When she discovered that her partner was also sexually abusing her young son, she left Central America and travelled to Mexico with two of her children.

In Mexico, MU's partner threatened to report her to the local authorities and to have her killed if she returned to her home country. MU felt that she had insufficient evidence to bring charges against her partner if she returned home, so she travelled to the United States to seek asylum. Under the United Na-

1) MD Canditate. Icahn School of Medicine at Mount Sinai.
2) Icahn School of Medicine at Mount Sinai. Correspondence to: jennifer.weintraub@mountsinai.org

* Non-peer-reviewed section. The views expressed in the contributions to this section are those of the authors and guest editors

https://doi.org/10.7146/torture.v32i3.132023

tion's Convention against Torture, MU meets criteria for torture as she experienced severe physical and emotional harm as a result of her gender and with acquiescence of the local police. She was evaluated shortly thereafter by a psychologist as part of her asylum case.

Ethical considerations

Written or verbal informed consent was obtained from the patient for de-identified information to be used for research and case reports.

Psychological signs and symptoms

MU was interviewed in her primary language of Spanish, via a professional interpreter. On examination, she was relaxed, cooperative, and articulate, with a normal thought process. Her affect was depressed and she became tearful when discussing painful and traumatic memories. She expressed hope for the future.

MU reported multiple psychological symptoms as a result of her trauma, including difficulty falling and staying asleep, frequent nightmares, and problems with concentration and memory. She described being easily startled and experiencing flashbacks of her trauma, especially when she saw accounts of intimate partner violence on television. Traumatic memories of her past frequently intruded her thoughts, causing her to cry and feel anxious. She also felt guilt and shame over what happened to her and her children.

During her evaluation, MU completed the Patient Health Questionnaire 9 (*PHQ-9*, a 9-item validated questionnaire for depression), and the Harvard Trauma Questionnaire (*HTQ*, a 16-item validated screening tool for Post-Trauma Stress Disorder in refugee populations). She completed a self-administered written Spanish language version of the PHQ-9 assessment on the computer. She scored a 4 (out of 27 possible points), which is below the cut-off for a diagnosis of depression. On the other hand, the HTQ was read aloud to MU in English by the evaluator, and then translated into Spanish by an interpreter. On this test, MU received a 2.52, with scores higher than 2.5 consistent with a diagnosis of PTSD.

Interpretation and conclusion

The psychological findings uncovered during MU's evaluation were highly consistent with the years of abuse MU described, and she met clinical criteria (per the American Psychiatric Association's *Diagnostic and Statistical Manual of Mental Disorders, known as the* DSM-V) for both Major Depressive Disorder and PTSD based on the symptoms she described. While she scored only a 4 on the PHQ-9, MU spoke of more severe depressive symptoms during her interview, including depressed mood, sleep difficulties, issues with concentration, and guilt. This discrepancy suggests that her low score on the PHQ-9 underestimated the severity of her depressive symptoms. Given that her father removed her from school in third grade, it was thought that her literacy level contributed to an inability of the self-administered questionnaire to measure her depressive symptoms.

In contrast, MU scored higher than 2.5 on the HTQ, exceeding the cut-off point for a diagnosis of PTSD. This diagnosis was consistent with her reported symptoms of intrusive thoughts, nightmares, flashbacks, hypervigilance, sleep disturbance, and difficulty concentrating. As the HTQ was administered verbally through a Spanish translator, it captured a more accurate assessment of her symptoms.

Discussion

In this case, the self-administered PHQ-9 vastly underestimated MU's current depressive symptoms, while the verbally adminis-

tered HTQ more accurately reflected her current psychological symptoms. Given the frequency that clinicians (especially those who are not mental-health specialists) utilise self-administered psychological assessments in forensic evaluations, the limitations of these tools must be considered and best practices established to avoid the pitfalls of their use.

The utility of diagnostic surveys may potentially be affected by the mode of questionnaire administration, as well as issues of literacy, language, and cultural difference. For example, face-to-face oral interviews can establish rapport and create space for open-ended questions. However, social desirability bias may cause the interviewee to water down responses to be more agreeable to the interviewer (Bowling 2005). Conversely, while self-administered surveys may suffer less from social desirability bias, they place a greater language and literacy burden on the survey taker (Bowling 2005), as in the case of MU.

Employing a mixed-mode design (e.g. interview followed by a self-administered survey) or alternative method of survey administration could help mitigate these biases (Bowling 2005) and limit the burden on the client. For example, telephone-administered PHQ-9s have been shown to yield similar results to self-administered PHQ-9s, demonstrating that a verbal administration reliably measures depression (Pinto-Meza et al. 2005). Additionally, some clinically validated surveys are available at different literacy levels. Assessing literacy prior to administration could also help to choose the right tool or delivery method (Olson et al. 2011). Finally, clinicians may consider the use of other scales to measure symptoms of depression, such as the Hamilton Depression Rating or Beck Depression Inventory. However, recent studies have suggested that the PHQ-9 is more accurate and reliable in distinguishing the severity of depression when compared with the Hamilton Depression Rating (Ma et al. 2021). When the PHQ-9 was compared with the Beck Depression Inventory (BDI-II), both scales were virtually interchangeable for assessing symptoms of depression. Given that the PHQ-9 is shorter and free to use as opposed to the copyrighted BDI-II, the study concluded that the PHQ-9 was still preferable for use (Kung et al. 2013). At many institutions, the PHQ-9 is the preferred questionnaire for depression screening during forensic asylum evaluations.

As the content of the PHQ-9 correlates directly with the DSM-V criteria for depression, it allows forensic evaluators to use the DSM-V as a standardized framework to easily support a diagnosis before a judge or legal system. It also provides structure for non-psychiatrist evaluators to supplement their physical exams with a brief, validated, psychiatric evaluation tool. That said, even with an improved mixed-method approach to survey administration, the cultural and linguistic limitations of questionnaires must also be considered. Studies have shown that the PHQ-9 is effective at assessing symptoms of depression across certain ethnic groups, cultures, and migration backgrounds, specifically showing validity and reliability in Spanish speaking countries in Central America such as Honduras (Wulson et al. 2002) and also Mexico (Arrieta et al. 2017), which can be applied to the case of MU. However, it remains possible that individual patients may face barriers to having symptoms of depression fully elicited by the PHQ-9 (Galemkamp et al. 2017; Zhong et al. 2014; Reich et al. 2018). First, for questionnaires created in English and translated into another language, the essence of a question may be lost in translation or not culturally understood in the way it was originally intended (Soukenik 2020). Moreover, conceptualizations of depression and mental health can vary across

cultures, which may pose a challenge when administering a questionnaire based on concepts derived from Western cultures (Lindheimer et. al 2020). Studies investigating how depression presents cross-culturally support the idea that different cultures may express and demonstrate various conditions differently (Hwang et al. 2008). There also may be significant stigma attached to expressing certain mood or other mental health symptoms, or perhaps a normalization of certain symptoms (fatigue, sleep difficulties) that a client may not identify as pathologic on a Likert-based scale like the PHQ-9. While there are likely universal forms of depressive symptoms that our tools can accurately identify, cultural variability should also be considered when working with a diverse patient population and making an accurate case for those we evaluate (Baradaran Eftekhari et al. 2021).

Overall, while validated psychological questionnaires are useful tools in forensic psychological evaluations, it is imperative to recognise their limitations and consider their results in conjunction with clinical judgement. In MU's case, an initial assessment of literacy level could have determined that an oral rather than written administration of the questionnaire would produce more accurate results. Still, even when such surveys are administered in a mode best for an individual client, significant obstacles may remain to applying them across cultures. In the case of MU, administering a verbal survey in her primary language yielded the most reliable description of her psychological symptoms, allowing the evaluator to accurately diagnose her and support the consistency of her experience.

References

Arrieta J., Aguerrebere M., Raviola G., Flores H., Elliott P., Espinosa A., Reyes A., Ortiz-Panozo E., Rodriguez-Gutierrez E. G., Mukherjee J., Palazuelos D., & Franke M. F. (2017). Validity and Utility of the Patient Health Questionnaire (PHQ)-2 and PHQ-9 for Screening and Diagnosis of Depression in Rural Chiapas, Mexico: A Cross-Sectional Study. *Journal of Clinical Psychology*, 73(9): 1076-1090. https://doi.org/10.1002/jclp.22390.

Baradaran Eftekhari, M., Mahmoodi, Z., Dejman, M., Forouzan, A. S., Falahat, K., Shati, M., Jandaghi, J., & Kane, J. (2021). Local perceptions of mental health in Iran, Semnan Province. *Brain and behavior*, *11*(2), e01971. https://doi.org/10.1002/brb3.1971

Bowling, A. (2005). Mode of Questionnaire Administration Can Have Serious Effects of Data Quality. *Journal of Public Health*, (3) 281-91. https://doi.org/10.1093/pubmed/fdi031.

Galenkamp, H., Stronks, K., Snijder, M.B., & Derks, E.M. (2017). Measurement invariance testing of the PHQ-9 in a multi-ethnic population in Europe: the HELIUS study. *BMC Psychiatry* 17, 349. https://doi.org/10.1186/s12888-017-1506-9

Hwang, W., Myers, H., Abekim, j. & Ting, J. (2008). A conceptual paradigm for understanding culture's impact on mental health: The cultural influences on Mental Health (CIMH) model. *Clinical Psychology Review*, *28*(2), 211–227. https://doi.org/10.1016/j.cpr.2007.05.001

Kung, S., Alarcon, R. D., Williams, M. D., Poppe, K. A., Jo Moore, M., & Frye, M. A. (2013). Comparing the Beck Depression Inventory-II (BDI-II) and Patient Health Questionnaire (PHQ-9) depression measures in an integrated mood disorders practice. *Journal of affective disorders*, *145*(3), 341–343. https://doi.org/10.1016/j.jad.2012.08.017

Lindheimer N., Karnouk C., Hahn E., Churbali, D., Schilz, L., Rayes, D., Balboui, M., & Boge, K. (2020). Exploring the Representation of Depressive Symptoms and the Influence of Stigma in Arabic-Speaking Refugee Outpatients. *Front Psychiatry*, 11:579057. https://doi.org/10.3389/fpsyt.2020.579057

Ma, S., Yang, J., Yang, B., Kang, L., Wang, P., Zhang, N., Wang, W., Zong, X., Wang, Y., Bai, H., Guo, Q., Yao, L., Fang, L., & Liu, Z. (2021). The patient health questionnaire-9 vs. the hamilton rating scale for depression in assessing major depressive disorder. *Frontiers in Psychiatry*, *12*. https://doi.org/10.3389/fpsyt.2021.747139

Olson, K, Smyth J.D., Wang, Y, & Pearson J.E. (2011). The Self-Assessed Literacy Index: Reliability and Validity. *Social Science Research*, 40(5), 1465-1476. https://doi.org/10.1016/j.ssresearch.2011.05.002.

Pinto-Meza, A, Serrano-Blanco, A, Penarrubia, M.T., Blanco, E. & Haro, J.M. (2005). Assessing Depression in Primary Care with

the PHQ-9: Can It Be Carried Out over the Telephone? *Journal of General Internal Medicine*, 20(8), 738-42. https://doi.org/10.1111/j.1525-1497.2005.0144.x.

Reich, H., Rief, W., Brähler, E., & Mewes, R. (2018). Cross-cultural validation of the German and Turkish versions of the PHQ-9: an IRT approach. *BMC Psychol* 6, 26. https://doi.org/10.1186/s40359-018-0238-z

Soukenik E, Haran H, Kirsch J, Pyakurel S, & Maleku A. (2021). Barriers and Facilitators of Mental Health Service Utilisation among Bhutanese Refugees in the USA: Findings from a Mixed-Methods Study. *Br J Soc Work*. https://doi.org/10.1093/bjsw/bcab123.

Wulsin L., Somoza E., & Heck J. (2002). The Feasibility of Using the Spanish PHQ-9 to Screen for Depression in Primary Care in Honduras. *Primary Care Companion Journal of Clinical Psychology,* 4(5): 191-195. https://doi.org/10.4088/pcc.v04n0504.

Zhong, Q., Gelaye, B., Fann, J. R., Sanchez, S. E., & Williams, M. A. (2014). Cross-cultural validity of the Spanish version of PHQ-9 among pregnant Peruvian women: A rasch item response theory analysis. *Journal of Affective Disorders*, *158*, 148–153. https://doi.org/10.1016/j.jad.2014.02.012