International Journal of Nursing Studies 56 (2016) 71–80

Contents lists available at ScienceDirect

International Journal of Nursing Studies

journal homepage: www.elsevier.com/ijns

Psychometric comparison of single-item, short, and comprehensive depression screening measures in Korean young adults

Hee-Ju Kim [a,*], Ivo Abraham [b,c,d]

[a] College of Nursing, Catholic University of Korea, Seoul, South Korea
[b] Department of Pharmacy Practice and Science, College of Pharmacy, University of Arizona, Tucson, AZ, USA
[c] Department of Family and Community Medicine, College of Medicine, University of Arizona, Tucson, AZ, USA
[d] Center for Health Outcomes and PharmacoEconomic Research, University of Arizona, Tucson, AZ, USA

ARTICLE INFO

Article history:
Received 22 May 2015
Received in revised form 11 December 2015
Accepted 12 December 2015

Keywords:
Metric systems
Validation studies
Depression
Psychometrics

ABSTRACT

Background: Integrating long depression-screening instruments into routine clinical practice and research studies is often impractical, necessitating short-item if not single-item measures with comparable psychometric properties.
Objective: To examine whether single-item or short depression-screening measures are comparable to a comprehensive screening measure in reliability (i.e., internal consistency and test–retest reliability) and validity (i.e., convergent, concurrent, and predictive validity) in Korean young adults within a Classical Testing Theory framework.
Method: A total of 458 students from six nursing colleges in South Korea completed three depression measures: the 20-item Center for Epidemiologic Studies-Depression screening instrument (CES-D; comprehensive measure); the five-item Profile of Mood States-Brief depression subscale (POMS-B depression subscale; short measure); a single-item Likert measure; and a single-item numeric rating scale. Internal consistency reliability was tested by Cronbach's alpha and item-total correlations; test–retest reliability by intraclass correlation coefficient (ICC); convergent validity by correlation with the CES-D; concurrent validity by the correlation with perceived stress level and sleep quality; and predictive validity by receiver operating characteristic curve to predict the two groups with different depression levels.
Results: The POMS-B depression subscale was comparable to the comprehensive CES-D scale in internal consistency reliability (alpha = .85); test–retest reliability (ICC = .76); and convergent (r = .81 with CES-D), concurrent (r = .64 with perceived stress level, r = .34 with sleep quality), and predictive validity (area under the curve = .88). The two single-item options were not comparable to the comprehensive CES-D.
Conclusion: The short POMS-B depression subscale shows an acceptable balance between practical clinical and research needs and psychometric quality.

What is already known about the topic?

- Screening for depression is critical in the care of young adults.
- The integration of lengthy screening instruments into routine clinical practice as well as research studies is impractical.

* Corresponding author at: Catholic University of Korea, College of Nursing, 505 Banpo Dong, Seo-Cho-Gu, Seoul, South Korea.
Tel.: +82 02 2258 7400; fax: +82 02 2258 7772.
E-mail address: heeju0906@gmail.com (H.-J. Kim).

http://dx.doi.org/10.1016/j.ijnurstu.2015.12.003
0020-7489/

- Comparative psychometric values for single-item or short measures are limited and further information is needed to enable the selection of shorter measures.

What this paper adds

- The five-item Profile of Mood States-Brief Depression subscale can be an appropriate, effective, and defensible alternative to a comprehensive depression screening measure.
- Single-item measures (a Likert-type and a numeric rating scale) showed much weaker reliability and validity than a 5-item short measure and a 20-item comprehensive measure.

1. Introduction

Depression is a worldwide public health issue. Up to half of the world's population may experience depression—subdiagnostic depressive symptomatology or manifest depressive illness—in their lifetime (Ferrari et al., 2013; Fushimi et al., 2013; Jonas et al., 2003). Worldwide, the major cause of death in young adults is suicide (Center for Disease Control, 2012; Korea National Statistical Office, 2013), most often associated with depression (Kang et al., 2015; Strandheim et al., 2014). Subdiagnostic depression at a young age can lead to severe mental health issues with poor long-term outcomes at a later age (Geller et al., 2001; Goldberg et al., 2001; Smith and Blackwood, 2004). Depression may interfere with one's social functioning and diminish one's sense of well-being (Herrman et al., 2002). Thus, screening for depression is critical in the care of young adults.

Researchers and clinicians have employed various types of depression-screening measures: single-item measures, short measures, and comprehensive measures. A comprehensive measure, for example, the Hamilton Depression Rating Scale (Hamilton, 1967), the Beck Depression Inventory (Beck et al., 1961), or the Center for Epidemiologic Studies-Depression (Radloff, 1977), consist of multiple items that measure detailed aspects of a selected concept. A short measure, for instance the World Health Organization (WHO) well-being index (1998) or the depression subscale of the Profile of Mood States-Brief (McNair et al., 1992), usually consists of five to 10 items and takes less than 5 min to complete. Due to the small number of items, a short measure may not be comprehensive enough to assess all aspects of the concept. A single-item measure is formatted as a single question, asking respondents to indicate their experience, emotions, or thoughts, using, for example, a Likert-type scale or a numeric rating scale.

Traditionally, researchers and clinicians have assumed that comprehensive measures of mental health conditions such as depression have better reliability and validity (Fried and Nesse, 2015; Gardner et al., 1998). Yet, respondent fatigue bias associated with answering highly similar questions can mitigate the effectiveness of comprehensive measures (Gardner et al., 1998; Robins et al., 2001). Also, summary scores from a comprehensive measure can be biased as they may include items that measure various levels of severity and cover multiple domains that incorporate different symptomatology phenotypes of depression (Cole et al., 2000; Fried and Nesse, 2015). That is, some items occur only in severely depressed subjects, whereas others occur in most subjects. Some subjects may have more psychomotor symptoms, whereas others may have more emotional symptoms. By adding scores from multiple items, information can be lost and bias introduced. In addition, comprehensive measures are lengthy and may be impractical in research or clinical contexts, where participants' burden and time constraints should be considered; thus, practitioners often prefer single-item or short measures (Bowling, 2005; De Boer et al., 2004; Mitchell and Coyne, 2007).

Ensuring the psychometric quality of short or single-item measures has been challenging. Patients with major depression (Zimmerman et al., 2006) reported a high association of single-item depression measures with a comprehensive measure and high reproducibility over a 1-week interval. However, a meta-analysis (Mitchell and Coyne, 2007) and a recent study with advanced cancer patients (Warmenhoven et al., 2012) showed lower or unacceptable sensitivity from a single-item measure of depression. Short measures, in contrast, have been shown to have better psychometric quality. For instance, a two-item depression measure (Arroll et al., 2005) and the WHO five-item well-being index (Henkel et al., 2003, 2004) showed acceptable sensitivity. The 4-item patient-health questionnaire (Löwe et al., 2010) showed acceptable internal consistency and validity. Still one can question whether one-item or short-measure instruments sufficiently represent complicated clinical presentations of depression. Also, it is unclear whether a certain type of single-item or short measure is more suitable when replacing a comprehensive measure. Most previous studies (a) mainly have examined validity, and particularly, sensitivity; (b) did not compare single-item measures, short measures, and comprehensive measures simultaneously; and (c) did not investigate whether single-item measures with different scaling methods (i.e., Likert vs. Numeric Rating Scale) differed in reliability and validity.

Using Classical Testing Theory as framework and designed accordingly, this study aimed to examine whether two single-item measures (i.e., a Likert-type and a numeric rating scale) and a short measure were comparable to a selected comprehensive depression-screening tool in reliability (i.e., internal consistency and test–retest reliability) and validity (i.e., convergent, concurrent, and predictive validity) in Korean young adults. We selected the 20-item Center for Epidemiologic Studies-Depression screening instrument (CES-D) as the comprehensive screening measure of depression, based on the following criteria: (a) it is a widely used, validated measure, (b) a validated Korean version exists, (c) it offers a clinical cut-off point to classify participants as potentially depressed, and (d) it is applicable to the general population. We selected the Profile of Mood States-Brief-Depression subscale (POMS-B depression subscale) as the short measure, based on the following criteria: (a) the number of items ranges from five to 10, (b) it is a widely used, validated measure, and (c) a Korean version exists.