

Original Article

Reliability testing of the HEARTSMAP psychosocial assessment tool for multidisciplinary use and in diverse emergency settings

Punit Virk BSc^{1,2}, Rob Stenstrom MD PhD^{1,3}, Quynh Doan MDCM, MHSc, PhD^{1,2,3}

¹School of Population and Public Health, Faculty of Medicine, University of British Columbia, Vancouver, British Columbia; ²BC Children's Hospital Research Institute, Vancouver, British Columbia; ³Department of Emergency Medicine, Faculty of Medicine, University of British Columbia, Vancouver, British Columbia

Correspondence: Quynh Doan, Faculty of Medicine, Department of Pediatrics, University of British Columbia, 4480 Oak Street, Vancouver, British Columbia V6T 1Z3. Telephone 604-875-2345 ext. 3691, fax 604-875-2366, e-mail qdoan@bcchr.ca

Abstract

Objective: HEARTSMAP is a tool developed to facilitate assessment and management of paediatric mental health (MH) patients by emergency department (ED) clinicians. We evaluate the inter-rater reliability of HEARTSMAP when administered by clinicians of various backgrounds.

Methods: In a cross-sectional study initiated in 2016, collaborating clinician evaluators (n=16) applied the HEARTSMAP tool to evaluate a set of 50 fictional clinical vignettes, digitally in an approach consistent with the anticipated tool's access and usage in clinical settings. Evaluators came from different types of health centres from across the province of British Columbia (Canada), including remote/rural, regional and urban academic health centres.

Results: We report moderate to near excellent agreement, overall among clinicians for all 10 of the tool's psychosocial sections ($\kappa=0.43$ to 0.93) and domain scores ($\kappa=0.75$ to 0.90), with acceptable agreement across all tool-triggered service recommendations ($\kappa=0.36$ to 0.65).

Conclusions: Our findings show that HEARTSMAP may be reliably used by ED clinicians in assessing MH issues among youth. Results from this study will assist in informing the wider clinical implementation of HEARTSMAP as a standard assessment tool, in diverse emergency care settings.

Keywords: *Emergency; Mental health; Pediatric; Assessment tool; Reliability; Multidisciplinary.*

Nearly one million Canadian youth live with mental health (MH) concerns. Given our fragmented MH system (1,2), emergency departments (EDs) have become 'safety nets' for families with MH concerns (2–4). Consequently, MH-related visits by youth and adolescents are rising across North American EDs, making up 7.2% of ED visits in the USA (5), and increasing at an annual rate of 3% to 7% in Canada (6–8). However, ED clinicians are often insufficiently trained in assessing level of MH risks and managing MH concerns (9). Standardized clinical MH tools have been recommended by the American Academy of Pediatrics Committee on Pediatric Emergency Medicine (10), however existing tools target specific concerns (e.g., depression, suicidality, anxiety) (11–13), are generally time- and energy-consuming for clinicians (14), or have not been validated for

ED use (15). Cappelli et al. developed HEADS-ED, a rapid MH screening tool, allowing clinicians to rate psychosocial issues' severity and determine if immediate intervention is needed (3). However, the current HEADS-ED does not clarify the urgency or types of services needed, and does not distinguish psychiatric from social or behavioral concerns (16).

To mitigate these challenges and respond to a growing reliance on ED clinicians to assess for psychosocial health and risk areas, an expanded but rapid psychosocial assessment and management tool called 'HEARTSMAP' (www.heartsmap.ca) was developed at the British Columbia Children's Hospital (BCCH) (17,18). HEARTSMAP allows clinicians to complete a comprehensive psychosocial assessment with a customized tool-generated report offering youth and

families specific service recommendations in addition to a suggested level of urgency. Unique to the tool is its ability to distinguish severity and acuity for psychiatric, social and behavioral issues.

The goal of our study was to evaluate the reliability of the HEARTSMAP tool in assessing child and youth MH complaints in a diverse range of EDs. We evaluate the inter-rater agreement on HEARTSMAP sectional and domain scoring patterns and tool triggered-recommendations, among ED clinicians from various types of medical care centres, ranging from small community-based, rural/remote, large regional and urban academic centres. Completion of this psychometric evaluation will ensure that the HEARTSMAP tool can be reliably utilized outside a paediatric quaternary care referral ED, by a diverse range of clinicians. High inter-professional reliability will ensure consistency in acute MH assessments in the ED, where physicians and various allied health workers are constantly in close interaction and collaboration to deliver integrated patient care.

METHODS

Study design and objectives

Between May and November 2016, we conducted a cross sectional study using a set of 50 fictional clinical vignettes to which study participants (n=16) applied the HEARTSMAP tool to perform a psychosocial assessment and trigger management recommendations. The primary objective of this study was to determine the inter-rater agreement in using the HEARTSMAP tool among a range of clinicians (emergency physicians, psychiatric liaison nurses, social worker and emergency nurses and nurse practitioners) who typically conduct emergency psychosocial assessments for youth.

Tool development

Since its development in 2014, HEARTSMAP has been piloted at the BCCH Pediatric ED for early implementation and acceptability testing. As an online algorithmic tool, HEARTSMAP facilitates clinicians in the collection of pertinent psychosocial information relating to 10 sections: *Home, Education and activities, Alcohol and drugs, Relationships and bullying, Thoughts and anxiety, Safety, sexual health, Mood and behaviour, Abuse and Professional resources*. Using section-specific guiding questions and scoring guidelines, clinicians can assess the severity of a patient's condition, on a scale 0 to 3, scoring a 0 (no concern), 1 (mild), 2 (moderate) or 3 (severe), for each of HEARTSMAP's 10 sections. Clinicians can also assess urgency of care, by recording whether resources pertaining to that area of concern have been accessed.

All sections on HEARTSMAP link to one of four possible domains: social, functional, youth health and psychiatry. Each domain is associated with recommendation for relevant services with several degrees of acuity of access based on its

score (composite of sectional scores) and what resources youth already have in place (for each section). Specific tool-recommendations include: psychiatric assessment, adolescent medicine or substance and addiction services, social work services to provide family support and educational support/counseling. Triggered recommendations are also urgency-specific ranging from emergent (in-hospital psychiatry consultation), urgent (community crisis response team with assessment within 72 to 96 hours), to less acute (outpatient speciality MH services, community-based MH clinics).

Study population and setting

Paediatric and community ED clinicians from four health authorities in British Columbia participated as collaborating evaluators in this study through email invitations and call outs by leadership members in emergency and MH services in involved health authorities. Fictional clinical vignettes covered a broad range of child and youth MH-related presentations and included information like age, gender and clinical information pertaining to the 10 section of the HEARTSMAP.

Data collection, storage, and analyses

Finalized clinical vignettes were stored on Research Electronic Data Capture (REDCap), a standardized online data collection tool (19). REDCap was also used to host the HEARTSMAP tool, which was enabled as an online survey that evaluators completed.

Data analyses were conducted using STATA 14.0 (Stata Corporation, College Station, Texas) and the Microsoft Excel 2010 Data Analysis Toolpak (Microsoft, Redmond, Washington) to conduct t-tests and determine P-values for professional group kappas. To assess inter-rater agreement during the training phase, evaluators' percent sectional agreement was computed and averaged for the three training cases. The small sample size did not allow for meaningful inter-rater reliability testing using kappa statistics. Quadratically weighted kappa statistics were computed to measure the inter-rater agreement among clinicians' domain and sectional scores for the 50 fictional vignettes. Simple Cohen's kappa statistics were calculated for tool-triggered recommendations. In accordance with Light (1971), kappa values were computed for all evaluator pairs, the overall mean of these pairs was then used as our index of agreement (20). Pairwise kappas were used to measure agreement specifically among physicians and nurses. Using weighted kappa statistics with quadratic weighting, allow consideration of the extent rater scoring disagreement, and offer comparability to intra-class correlation coefficients (ICC) (21). Subgroup analysis evaluating clinicians' inter-rater agreement by sets of 10 vignettes was conducted, to assess any patterns that emerged in weighted kappa values over the 50 cases. We estimated that 800 evaluated vignettes were required to achieve with

80% power, 60% agreement (correlating with a moderate agreement Kappa range) with 5% precision. With each clinician evaluating 50 vignettes, 16 clinicians would meet our required sample size (22).

Study procedure

Participating evaluators were trained to comprehensively understand and effectively use the tool in the clinical setting. The overall aim of these training sessions was to ensure that HEARTSMAP scoring and associated management recommendations were approached in a manner consistent with tool access and use in the clinical setting when HEARTSMAP is implemented. Training consisted of watching a 25-minute instructional video, followed by independent assessment of three test cases using HEARTSMAP, which were then reviewed with the main investigators during a 20-minute telephone conversation to identify sources of discrepancies in scoring interpretation and discuss areas requiring further refinement or explanation, no evaluator received any additional training to what is described above. Subsequently, each evaluator received a link via email to access the clinical vignettes and the MyHEARTSMAP tool. Upon completing their assessment of vignettes, scores and triggered recommendations were recorded automatically to the study's REDCap database for analyses.

RESULTS

A total of 16 paediatric and community ED clinicians were recruited and used the HEARTSMAP tool to evaluate the 50 clinical vignettes. Approximately equal proportions of clinicians were recruited from across the four major health authorities in British Columbia. Additional details on these evaluators are summarized in Table 1.

For the three training cases, average clinician percent agreement ranged from 61.9% (Education and activities) to 100% (Alcohol and drugs). Across HEARTSMAP's 10 sections, we observed weighted kappas ranging from 0.46 (Professional and services) to 0.93 (Alcohol and drugs). Mean weighted kappas for all sections and sub-groups analyses are reported in Table 2. At the domain-level (Table 3), we observed substantial

to almost perfect inter-rater scoring agreement, with weighted kappa statistics of 0.75 (0.73 to 0.76) for psychiatry, 0.84 (0.83 to 0.85) for youth health, 0.89 (0.89 to 0.90) for social and 0.78 (0.77 to 0.79) for the functional domain. Across all clinicians, subgroup analysis looking at inter-rater agreement on sets of 10 vignettes did not show noticeable differences in agreement over the progression of the 50 cases.

Simple kappa statistics were computed for HEARTSMAP services recommendation, triggered on the basis of a clinician's sectional and domain scores, as well as whether they assess youth as having existing appropriate resources in place for a particular section. These management options show inter-rater agreement of 0.43 (0.40 to 0.46) for psychiatric consultation, 0.46 (0.43 to 0.50) for crisis response, 0.36 (0.33 to 0.39) for community-based MH clinics, 0.52 (0.49 to 0.55) for youth health clinics and 0.65 (0.62 to 0.68) for social work services. Mean inter-rater agreement on sectional scores, domain scores, and triggered-recommendations, among clinician groups and comparisons across subgroups are reported with 95% confidence intervals in Table 4.

Agreement among physicians was significantly higher on 'Alcohol & drugs' and 'Abuse' sections, compared to allied health workers, and similarly on 'Social' domain scores ($P < 0.001$). Whereas allied health workers showed higher kappas for 'Thoughts & anxiety', 'Sexual health' and 'Mood & behavior' ($P < 0.05$). While allied health workers showed higher kappas for psychiatric domain scores ($P = 0.026$), agreement on most psychiatric tool-triggered recommendations (community-based MH clinics, crisis response intervention) was higher among physicians ($P < 0.05$), as well as on referral to social work services ($P < 0.001$).

DISCUSSION

HEARTSMAP is one of few known clinician-administered, digital and broad-based psychosocial assessment tools for youth, designed exclusively for the ED. Here, we conducted an evaluation of the tool's inter-rater agreement when applied by a diverse sample of physicians and allied health workers to a set of fictional vignettes of varying ED MH-related presentations.

Table 1. Distribution of collaborating clinician evaluators by discipline and hospital types

Centres	MDs (N=7)			Allied Health (N=9)	
	EMs*	PEMs**	Psychiatrists	Nurses	Social Workers
Paediatric referral centre	0	2	1	3	0
Urban community hospital	3	0	0	2	2
Remote regional hospital	1	0	0	1	1
N (%) Total=16	4	2	1	6	3

*Emergency Physicians

**Paediatric Emergency Physicians

Table 2. Quadratically weighted kappa statistics (95% confidence intervals) for sectional score agreement among evaluating clinicians

Section	All Clinicians	Physicians Only	Allied Health Only	P-values* (two-tail)
Home	0.78 (0.77–0.80)	0.78 (0.76–0.80)	0.78 (0.75–0.81)	>0.1
Education and activities	0.70 (0.69–0.70)	0.70 (0.66–0.74)	0.69 (0.66–0.74)	>0.1
Alcohol and drugs	0.93 (0.93–0.94)	0.96 (0.95–0.97)	0.91 (0.90–0.93)	<0.001
Relationships and bullying	0.68 (0.66–0.70)	0.64 (0.58–0.70)	0.71 (0.68–0.74)	0.05
Thoughts and anxiety	0.91 (0.90–0.91)	0.89 (0.87–0.91)	0.92 (0.90–0.93)	0.025
Safety	0.82 (0.81–0.83)	0.80 (0.77–0.84)	0.82 (0.80–0.84)	>0.1
Sexual health	0.90 (0.89–0.91)	0.85 (0.82–0.89)	0.94 (0.93–0.96)	<0.001
Mood and behavior	0.69 (0.68–0.71)	0.68 (0.64–0.71)	0.74 (0.71–0.76)	0.009
Abuse	0.88 (0.87–0.90)	0.93 (0.90–0.95)	0.84 (0.81–0.88)	<0.001
Professionals and services	0.43 (0.40–0.46)	0.38 (0.28–0.47)	0.46 (0.41–0.51)	>0.1

*Comparing agreement among physicians versus agreement among allied health professionals.

Table 3. Quadratically weighted kappa statistics (95% confidence intervals) for domain score agreement among various ED clinician raters

Domain	All Clinicians	Physicians Only	Allied Health Only	P-values* (two-tail)
Social	0.90 (0.89–0.90)	0.92 (0.91–0.93)	0.88 (0.87–0.89)	<0.001
Functional	0.78 (0.77–0.79)	0.79 (0.75–0.82)	0.78 (0.76–0.79)	>0.1
Youth Health	0.84 (0.83–0.85)	0.83 (0.80–0.86)	0.85 (0.84–0.86)	>0.1
Psychiatry	0.75 (0.73–0.76)	0.72 (0.68–0.76)	0.78 (0.75–0.81)	0.026

*Comparing agreement among physicians versus agreement among allied health professionals.

ED emergency department

Table 4. Simple kappa statistics for interrater agreement of various ED clinicians on tool triggered recommendations with 95% confidence intervals.

Recommendations	All Clinicians	Physicians Only	Allied Health Only	P-values* (two-tail)
Social support services	0.65 (0.62–0.68)	0.73 (0.68–0.77)	0.58 (0.52–0.64)	<0.001
Adolescent/Substance & addiction services	0.52 (0.49–0.55)	0.57 (0.48–0.65)	0.48 (0.44–0.52)	>0.1
Community based mental health services	0.36 (0.33–0.39)	0.48 (0.39–0.56)	0.29 (0.25–0.33)	<0.001
Crisis Response Team (urgent community based mental health services)	0.46 (0.43–0.49)	0.55 (0.45–0.64)	0.41 (0.37–0.46)	0.016
Psychiatry (in-house urgent psychiatric evaluation and interventions)	0.43 (0.40–0.46)	0.49 (0.41–0.57)	0.40 (0.35–0.44)	0.063

*Comparing agreement among physicians versus agreement among allied health professionals.

ED emergency department

Promising evidence was demonstrated for the tool's inter-rater reliability, with substantial and moderate, to almost perfect sectional and domain scoring agreement across all evaluating clinicians, and fair or above on all tool-triggered recommendations. As HEARTSMAP is grounded within the framework of communitric theory, a Cronbach alpha measure may not be appropriate to evaluate internal consistency (23). However, high agreement among evaluators, across the four domain scores is indicative of internal consistency with the tool's scoring logic.

Nonchance agreement and disagreement in clinician scoring was assessed using quadratically weighted kappa statistics. No

significant differences were found for inter-professional group agreement on 'Home', 'Education & activities', 'Safety' and 'Professional & resources' sections, however physicians showed higher agreement for 'Alcohol & drugs' and 'Abuse'. Allied health workers displayed greater agreement on 'Relationships & bullying', 'Thoughts & anxiety', 'Sexual health' and 'Mood & behavior' sections. In a population of adolescents with recent cancer diagnoses, Hedström et al. found nurses showing higher sensitivity for distress related to psychosocial factors and physicians more sensitive toward physical distress, using the hospital anxiety and depression scale (HADS) (24). Findings seen here may help to

explain the inter-rater agreement patterns we observed in our study. Physicians may focus on the absence or presence of treatment-related problems such as alcohol and drugs or the physical signs of abuse, and more consistently score the severity of such concerns. While nurses and other allied health workers may be more attuned to the patients' general quality of life such as their mood, relationships and thoughts, subsequently assessing any concerns in these areas more consistently.

Other paediatric MH assessment tools have reported good inter-rater reliability among clinicians from varying disciplines, such as the Health of the Nation Outcome Scales for Children and Adolescents (HoNOSCA), Global Assessment of Psychosocial Disability (GAPD), and the Children's Global Assessment Scale (CGAS) (25). However, these tools have not been evaluated in emergency settings and do not provide clinicians with emergency disposition guidance. The closest instrument to HEARTSMAP in this regard is HEADS-ED, the only other known tool designed specifically to assess youth in the ED. Cappelli et al. compared inter-professional agreement between crisis intervention workers and paediatric emergency physicians in screening a sample of 140 patients using HEADS-ED (26). Scores on HEADS-ED lead to a binary outcome: requires further MH assessments and interventions, or not. For items related to psychiatric disposition from the ED, reported ICCs were 0.529 for Suicidality, 0.208 for Emotions and behaviors and 0.292 for Discharge resources. Our measure of agreement for the HEARTSMAP sections which are represented in HEADS-ED were generally higher, which may be attributed to our differing methodological approach. In the HEADS-ED study, clinicians applied the instrument to youth presenting in the ED with MH-related complaints. However, studies have found that using clinical vignettes provide similar results to that of a standardized patient, and clinician responses to vignettes generally reflect their response to real world interactions (27–29). We also report the participation of a diverse sample of allied health workers and physicians (Table 1), who commonly interact and work collaboratively to provide integrative medical care in the ED.

Our study was not without limitations. While we could not identify a learning curve in clinicians' application of HEARTSMAP to a set of clinical vignettes, this may have been due to a lack of control over the order and how diligently or distractedly they approach the cases. Additionally, use of vignettes prevents reviewers from seeking clarification of clinical details and from testing the information gathering component of a clinical encounter. Among the 16 participating clinicians, we also found that two evaluators showed lower agreement across the five tool recommendations. Both outliers were ED physicians, with one spending nearly half the amount of time (5 hours) evaluating the 50 vignettes, compared to a median of 10 hours spent by other clinicians. However, exploratory sensitivity analysis excluding these

outliers confirmed that the level of agreement on scoring and recommendations did not change. With wider tool implementation in British Columbia EDs and utilization by more clinicians, we believe that its reliability and validity may not be greatly affected by outliers, which can be expected to arise in the larger ED clinician population. While we observed acceptable inter-rater reliability on tool-triggered recommendation, entries were not forced for sections indicating existing services in place. As there were instances where clinicians left this section empty, this would affect agreement on recommendations, as depending on how clinicians scored this section different service referrals would get triggered. Moreover, services in-place options for clinicians to choose from in this section were generic and lacking descriptors, which might have introduced additional challenges to the evaluators. With HEARTSMAP's on-going implementation in community EDs across British Columbia, we hope to avoid these issues by having modified the tool such that clinicians are forced to answer to whether more specific services are already in place for patients, on specific tool-sections. While we have future plans to build a mobile version of the tool and expand access, HEARTSMAP is currently only available online and limited to clinicians in British Columbia.

CONCLUSION

HEARTSMAP shows promise as a reliable instrument in offering standardized and broad-based MH assessment which can be used by a diverse range of ED clinician types.

This study was reviewed and approved by the University of British Columbia Children's and Women's Hospital ethics review board.

Acknowledgments

We would like to acknowledge the participation of the following clinicians who evaluated fictional vignettes using HEARTSMAP: Dr Tyler Black, Lisa Buono, Raj Chalal, Zoe Chee, Mandy Deresh, Debra Forcier-Salverda, Susie Girling, Dr Cecilia Kim, Dr Celine Kim, Dr Justin Kopp, Dr Julian Marsden, Visnja Mitrovic, Dr Badrinath Narayan, Craig Norris and Carmella Silvestri.

Funding

QD reports grants from Specialist Service Committee, a collaboration between the Doctors of British Columbia association and the British Columbia Ministry of Health, and a salary/scholar's award paid to UBC for research protected-time from Michael Smith Foundation for Health Research during the conduct of the study.

Conflict of Interest

The authors have no conflict of interest to declare.

References

1. Bartram M, Chodos H. Changing directions, changing lives: The mental health strategy for Canada. *Can J Community Ment Heal* 2013;32(4):1–8.
2. Waddell C, McEwan K, Shepherd CA, Offord DR, Hua JM. A public health strategy to improve the mental health of Canadian children. *Can J Psychiatry* 2005;50(4):226–33.
3. Cappelli M, Gray C, Zemek R, et al. The HEADS-ED: A rapid mental health screening tool for pediatric patients in the emergency department. *Pediatrics* 2012;130(2):e321–7.
4. Baren JM, Mace SE, Hendry PL, Dietrich AM, Goldman RD, Warden CR. Children's mental health emergencies - Part 1 challenges in care: Definition of the problem, barriers to care, screening, advocacy, and resources. *Pediatr Emerg Care* 2008;24(6):399–408.
5. Simon AE, Schoendorf KC. Emergency department visits for mental health conditions among US children, 2001–2011. *Clin Pediatr (Phila)* 2014;53(14):1359–66.
6. Newton AS, Ali S, Johnson DW, et al. A 4-year review of pediatric mental health emergencies in Alberta. *CJEM* 2009;11(5):447–54.
7. Mapelli E, Black T, Doan Q. Trends in pediatric emergency department utilization for mental health-related visits. *J Pediatr* 2015;905–10.
8. Canadian Institute for Health Information. Care for Children and Youth With Mental Disorders 2015. Available from: https://secure.cihi.ca/free_products/CIHI_CYMH_Final_for_pubs_EN_web.pdf.
9. Zun L. Care of psychiatric patients : The challenge to emergency physicians. *West J Emerg Med* 2016;17(2):173–6.
10. Dolan MA, Fein JJA, The Committee on pediatric emergency medicine. Technical report - pediatric and adolescent mental health emergencies in the emergency medical services system. *Pediatrics*. 2011;127(5):e1356–66.
11. Horowitz LM, Wang PS, Koocher GP, et al. Development of a brief screening tool. *Pediatrics* 2001;107:1133–7.
12. Ramsawh HJ, Chavira DA, Kanegaye JT, Ancoli-Israel S, Madati PJ, Stein MB. Screening for adolescent anxiety disorders in a pediatric emergency department. *Pediatr Emerg Care* 2012;28(10):1041–7.
13. Rutman MS, Shenassa E, Becker BM. Brief screening for adolescent depressive symptoms in the emergency department. *Acad Emerg Med* 2008 16;15(1):17–22.
14. Vu F, Daeppen J, Hugli O, et al. Screening of mental health and substance users in frequent users of a general swiss emergency department. *BMC Emerg Med* 2015;15(1):27.
15. Chun T, Duffy S, Linakis J. Emergency department screening for adolescent mental health disorders: The who, what, where, why and how it could and should be done. *Clin Pediatr Emerg Med* 2013;14(1):199–216.
16. Cappelli M. The HEADS-ED: Review of a mental health screening tool for pediatric patients. *Psychiatric Times* 2012 (cited July 18, 2016). p. 1. Available from: <http://www.psychiatrictimes.com/child-adolescent-psychiatry/heads-ed-review-mental-health-screening-tool-pediatric-patients>.
17. Doan Q, Black T. HEARTSMAP 2015 (cited September 12, 2016). Available from: <http://heartsmmap.ca/>.
18. Lee A, Black T, Meckler G, Eslami A, Doan Q. Psychometric Properties of HEARTSMAP, a Psychosocial Assessment Tool Applied to Children and Youth with a Mental Health Elated Pediatric Emergency Visit. In: 28th International Congress of Pediatrics. Vancouver; 2016. p. 282.
19. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap) - A metadata driven methodology and workflow process for providing translational research informatic support. *J Biomed Inform* 2009;42(2):377–81.
20. Light RJ. Measures of response agreement for qualitative data: Some generalizations and alternatives. *Psychol Bull* 1971;76(5):365–77.
21. Hallgren KA. Computing inter-rater reliability for observational data: An overview and tutorial. *Tutor Quant Methods Psychol* 2012;8(1):23–34.
22. Hong H, Choi Y, Hahn S, Park SK, Park BJ. Nomogram for sample size calculation on a straightforward basis for the kappa statistic. *Ann Epidemiol* 2014;24(9):673–80.
23. Lyons JS. Measurement as communication. In: *Communimetrics*. New York, NY: Springer US, 2009:19–44. Available from: http://link.springer.com/10.1007/978-0-387-92822-7_2.
24. Hedström M, Kreuger A, Ljungman G, Nygren P, von Essen L. Accuracy of assessment of distress, anxiety, and depression by physicians and nurses in adolescents recently diagnosed with cancer. *Pediatr Blood Cancer*. 2006;46(7):773–9.
25. Hanssen-Bauer K, Gowers S, Aalen OO, et al. Cross-National reliability of clinician-rated outcome measures in child and adolescent mental health services. *Adm Policy Ment Heal Ment Heal Serv Res*. 2007;34(6):513–8.
26. Cappelli M, Zemek R, Polihronis C, et al. The HEADS-ED Evaluating the clinical use of a brief, action-oriented, pediatric mental health screening tool. *Pediatr Emerg Care* 2017. doi:10.1097/PEC.0000000000001180. [Epub ahead of print]
27. Hanssen-Bauer K, Gowers S, Aalen OO, et al. Cross-national reliability of clinician-rated outcome measures in child and adolescent mental health services. *Adm Policy Ment Health* 2007;34(6):513–8.
28. Peabody JW, Luck J, Glassman P, et al. Measuring the quality of physician practice by using clinical vignettes: A prospective validation study. *Ann Intern Med* 2004;141(10):771–80.
29. Veloski J, Tai S, Evans AS, Nash DB. Clinical vignette-based surveys: A tool for assessing physician practice variation. *Am J Med Qual* 2005;20(3):151–7.