

Reliability and Validity Testing of Pilot Data from the TeamSTEPPS® Performance Observation Tool

Mary Beth Maguire*, Marie N Bremner and Daniel J. Yanosky

Kennesaw State University, Kennesaw, United States.

*Corresponding author : Mary Beth Maguire, Kennesaw State University, Kennesaw, United States, E-mail: mmaguir5@kennesaw.edu

Received date: August 25, 2014, Accepted date: September 18, 2014, Published date: September 22, 2014

Copyright: © 2014 Maguire MB, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Abstract

Background: The TeamSTEPPS® Performance Observation Tool (TPOT) is an instrument used in the evaluation of team performance: however, no assessment of the tool's reliability or validity exists among nurse educators.

Methods: A convenience sample of 31 nurse educators completed the TPOT to assess the reliability and validity of the instrument.

Results: Using Cronbach's alpha, the TPOT demonstrated a strong internal consistency coefficient. Through cross-group analysis of scoring between undergraduate and graduate nursing faculty, some evidence for convergent validity was confirmed.

Conclusion: This pilot study establishes the internal consistency reliability and convergent validity of the TPOT instrument when used by nurse faculty.

Keywords TeamSTEPPS® performance observation tool; Simulation team training; Reliability

Reliability and Validity Testing of Pilot Data from the TeamSTEPPS® Performance Observation Tool

Fifteen years have passed since the Institute of Medicine first published *To Err is Human* [1]. This groundbreaking work described hospital errors as the eighth leading cause of death among patients in the United States. Many of the hospital errors identified were not a result of technical incompetence but rather human factors. Today, such medical errors persist and as recently as 2010, 180,000 deaths were attributed to them [2].

Background

TeamSTEPPS®

One way to improve human factors and minimize medical errors is through the TeamSTEPPS® curriculum. TeamSTEPPS® stands for Team Strategies and Tools to Enhance Performance and Patient Safety. TeamSTEPPS® is a comprehensive set of materials and training curriculum which seeks to improve patient safety through the use of team-based principles. The TeamSTEPPS® program was created by the Agency for Healthcare Research and Quality (AHRQ) and the Department of Defense (DOD). The curriculum is an evidence-based program based on 25 years of research related to teamwork, team training, and culture change [3]. TeamSTEPPS® was adopted as the national standard for healthcare team training in November 2006 [4]. However, despite TeamSTEPPS® set as the gold standard for health care team training, to date there has not been testing for reliability and

validity of a tool that measures team performance based upon the curriculum.

The TeamSTEPPS® program is comprised of four primary teamwork skills: leadership, communication, situation monitoring, and mutual support. The TeamSTEPPS® curriculum reinforces the use of behaviors such as Situation-Background-Assessment-Recommendation (SBAR), check-back, and huddle which seek to improve team performance [3]. The implementation of TeamSTEPPS® principles has proven to reduce negative patient outcomes [5]. One hospital reports a 30% reduction in medical errors and an 88% decrease in the number of patient falls after implementing TeamSTEPPS® training [6].

The simulated clinical experience provides an ideal opportunity for learners to practice and refine clinical skills, teamwork, and communication in a controlled environment under the direction of faculty seeking to achieve a set of pre-determined objectives [7]. Combining simulation and the TeamSTEPPS® curriculum is an effective teaching strategy to allow learners the opportunity to engage in experiences addressing knowledge, skill, and interpersonal interactions while practicing team strategies in a safe and reproducible environment.

Simulation Instruments

An important aspect of determining the effectiveness of simulated experiences is through evaluation. Most instruments used to measure student performance lack reported reliability and validity [8]. The utilization of instruments that have undergone appropriate psychometric testing is necessary to support reliable and valid assessment of student performance. Howard found a large number of untested instruments in current use and suggested a moratorium on

further simulation instrument development until the appropriate psychometric assessments have been completed with instruments currently available [9]. In order to advance the simulation pedagogy, performance evaluation instruments from an individual and team perspective must undergo the rigor of psychometric testing.

The TeamSTEPPS® Performance Observation Tool

The TeamSTEPPS® Performance Observation Tool (TPOT) is a 25 item instrument used to evaluate 5 domains of team performance. The domains are: team structure, leadership, situation monitoring, mutual support, and communication (Figure 1). The TPOT uses a 5-point scale that ranges from 1 (very poor) to 5 (excellent). The maximum score possible on the TPOT is 125 points.

The AHRQ and the DOD created the TPOT in an effort to quantify team performance. The tool creators acknowledge the TPOT had not been tested for reliability or validity prior to its publication nor is a standardized user menu or scoring method available [10]. At the time of this publication, no report of reliability or validity of the use of the TPOT among nurse educators was found in the literature.

Purpose

The purpose of this study was to assess the internal consistency reliability and convergent validity of data produced by the TPOT when used by a group of nurse educators at one university in a southeastern state for evaluation of third semester Baccalaureate nursing students' team performance during a post-partum hemorrhage simulated patient care scenario. The research question was: What is the internal consistency reliability and convergent validity of TPOT

Methods

Convenience sampling was utilized to recruit study participants. All full-time faculty and one cohort of nursing doctoral students from one school of nursing in a southeastern state were recruited. The inclusion criteria were for participants to be over 18 years of age and currently employed as a nurse educator. Institution Review Board approval was requested and exempt status was achieved.

Data Collection

Data were collected in individual sessions or group sessions to accommodate the schedules of participants. Group session participants were instructed to avoid engaging in verbal and non-verbal communication to avoid scoring bias. Demographic information of participants was obtained. Participants viewed a 10-minute pre-recorded clinical simulation scenario of third semester Baccalaureate nursing students caring for a patient experiencing a post-partum hemorrhage. Participants viewed the scenario two consecutive times. The first viewing was to observe the overall scenario content. After the initial viewing and prior to the second scenario viewing participants reviewed the TPOT instrument and received scripted scoring instructions. The second viewing occurred immediately after the first to allow scoring of the TPOT. Participants were granted no more than 10 minutes of additional time at the end of the second viewing to complete TPOT scoring.

Data Analysis

Data were analyzed using PASW® Statistics GradPack 18 for Mac® and the SAS V9.3 system. Univariate analysis was used to examine the demographic nature of the sample. Internal consistency is the reliability estimate of a test based on a single administration [11]. To provide an estimate of internal consistency reliability, Cronbach's alpha coefficient was used along with split-half analysis. The split-half correlation is an additional method of analysis to further suggest the reliability of data as the TPOT was administered to participants on one occasion [12]. Convergent validity was assessed using 1-way analysis of covariance (ANCOVA) to detect possible differences in TPOT total scores among two distinct groups: those who teach undergraduate vs. graduate nursing courses while controlling for number of years of experience (Table 1). The TPOT total scores should ideally reflect only the quality of the scenario being evaluated and minimize subjectivity from the rater. Therefore, to establish convergent validity it was hypothesized that no systematic differences would exist among raters based on years of experience and level of teaching responsibilities. In other words, it was hypothesized that everyone in this sample would possess similar skills to evaluate the scenario at hand and therefore the number of years of experience and the level of teaching responsibilities would not systematically impact their TPOT total scores.

Results

Sample

Thirty-one participants were enrolled in the study (Table 2). Education preparation of the group approached equal balance between faculty with doctoral degrees (52%) and faculty with Master's degrees (48%). Teaching responsibilities among the group were equally distributed between baccalaureate (52%) and graduate (48%) degree levels. More than half of participants had completed five or more education courses at the graduate level (55%). The majority of the group had been teaching five or more years (68%). The group was nearly balanced between those currently responsible for performance based testing (48%), and those with no experience or no current experience with performance based testing (52%). A small percentage of participants indicated active teaching of the TeamSTEPPS® curriculum or of being a certified TeamSTEPPS® trainer

Reliability

As a measure of internal consistency, Cronbach's alpha coefficient was 0.965. Split-half analysis of the TPOT was 0.943 (13 items) and 0.952 (12 items). The overall mean score of the TPOT was 70.77 (SD=21.42) (Table 1). Given all of these pieces of evidence, we find the internal consistency reliability of the TPOT to be strong.

Validity

Instrument validity refers to how well an instrument actually measures what it is supposed to measure (Field, 2009). As a measure of convergent validity (one type of instrument validity), TPOT total scores were not found to differ significantly based on level of teaching responsibility and years of experience, $F_{(1,29)} = 0.26$, $p = .6107$, $\omega^2 = 0.04$. Once again, these results were expected as no systematic differences here suggest that raters are in fact evaluating the same latent construct regardless of the raters' years of experience and level of teaching responsibilities. Graduate faculty total TPOT scores were somewhat lower ($M=64.07$, $SD=19.0$) than undergraduate faculty

(M=77.06, SD=22.21), however, these differences were not found to be statistically significant (Table 2). Subsequent ANCOVA analyses were performed on the 25 individual items of the TPOT as well. Results showed no statistical significance at the $\alpha = .05$ level.

Figure 1: Team Performance observation Tool.

Variable	Category	N (%)
Type of program taught	Baccalaureate	16 (52%)
	Graduate	15 (48%)
Number of years as nursing faculty	<1	1 (3%)
	1-3 year	6 (19%)
	4-5 years	3 (10%)
	6-10 years	4 (13%)
	>10 years	17 (55%)
Highest degree completed	Master's Degree	15 (48%)
	Doctoral Degree	16 (52%)
Number of education courses completed at the graduate level	0	4 (13%)
	1	1 (3%)
	2	4 (13%)
	3	1 (3%)
	4	4 (13%)

	5 >5	2 (7%) 15 (48%)
Experience with clinical simulation	No experience Participated as a learner in a simulated clinical environment Facilitated learning in a clinical environment Attended a simulation workshop Took a simulation elective course	7 (23%) 5 (16%) 17 (55%) 1 (3%) 1 (3%)
Experience with evaluating students in a performance based assessment	No experience Some experience but none currently Responsible for evaluating performance based learning this semester	4 (13%) 12 (39%) 15 (48%)
Indicate your experience with team work and the TeamSTEPS curriculum	Experience with teamwork concepts, not TeamSTEPS Familiar with TeamSTEPS curriculum Teach TeamSTEPS A certified TeamSTEPS trainer	15 (48%) 13 (42%) 1 (3%) 2 (7%)

Table 1: Participants Demographics (N=31)

Attribute/Group	N	Mean	Median	Std Dev	Minimum	Maximum	Lower 95% Confidence Limit	Upper 95% Confidence Limit
Years of Experience	31	3.97	5.00	1.33	1.00	5.00	3.48	4.46
TPOT Total Score	31	70.77	66.00	21.42	28.00	119.00	62.92	78.63
TPOT Total Score by Group								
Undergraduate	16	77.06	73.50	22.21	43	119	65.23	88.90
Graduate	15	64.07	63.00	19.00	28	97	53.55	74.59

Table 2 : TPOT Question Scores for Undergraduate and Graduate Faculty 1-way ANOVA (Undergraduate N=16 Graduate N=15)

Discussion

This study serves to establish psychometric properties related to reliability and validity of TPOT pilot data when used to evaluate team skills of undergraduate nursing students. The findings of this study suggest the TPOT is a reliable instrument to utilize in the evaluation of team performance among undergraduate nursing students. Furthermore, this study provides the beginnings of a validity study through presenting initial evidence of convergent validity. Specifically,

findings suggest that convergent validity is acceptable with no statistical differences detected among teaching responsibility groups while controlling for years of experience. T

The results of TPOT data can yield worthwhile data to healthcare teams seeking to evaluate team performance. Nurse educators may find the TPOT an effective tool to utilize in the simulated clinical environment to provide formative and/or summative assessment of team performance. Data from TPOT scoring can also provide information to educators as to the effectiveness of leadership and communication within a curriculum.

Limitations

Several limitations of this study must be noted. First, variability occurred in how data was collected. Some participants completed the TPOT in one-on-one sessions and others completed the TPOT in small group sessions all led by the principle investigator. These small group sessions may have resulted in verbal and non-verbal communication among participants that influenced participants' TPOT score however, an attempt was made to control and minimize any such bias. Second, the scenario reviewed was a performance of novice practitioners. Various skill imperfections were present and may have distracted participants from the overall team performance. Third, the sample size was small, not randomized, and limited to one institution. Additional testing is required with a larger sample size to complete a factor analysis and compare findings among multiple institutions. While all faculty were experienced with broad concepts of teamwork, not all faculty members were actively teaching TeamSTEPPS® curriculum nor were they certified TeamSTEPPS® trainers. Unfamiliarity with the TeamSTEPPS® curriculum may have accounted for the variability in scoring. Thus, it is recommended future studies mandate certification in TeamSTEPPS® as part of the inclusive criteria for participants.

Conclusion

Clinical simulation provides an ideal setting for providing reliable and valid assessment of student performance across multiple domains. Recommendations for further studies include repeated measures of TPOT scoring among groups to evaluate the stability of the instrument and utilizing the TPOT with simulation scenarios that differ in level of

team performance to determine if the tool can detect varying abilities of team performance. Increasing the sample size of participants will strengthen the precision of psychometric indicators of the TPOT. Research is in progress on a national scale to replicate this study with other schools of nursing faculty to determine the reliability and validity of TPOT data among a larger sample

Refined instruments for the evaluation of teams will help to standardize assessment of team performance in the simulated clinical environment and ultimately the clinical practice setting. The improved performance of healthcare teams in clinical practice will help to mitigate human factors and result in reduction of medical errors.

References

1. Institute of Medicine (1999) To err is human: Building a safer health system.
2. Medical errors in the USA: Human or systemic (2011) *The Lancet*.
3. Agency for Healthcare Research and Quality (2013) TeamSTEPPS: National implementation.
4. American Institutes for Research (2010) TeamSTEPPS Teamwork Perception Questionnaire (TTPQ)
5. Courtright SH, Steward GL, Ward MM (2012) Applying research to save lives: Learning from team training approaches in aviation and health care. *Organizational Dynamics*, 41: 291-301.
6. Nailberk D (2012) Hospital cuts med errors 30%, falls 88% with TeamSTEPPS. *Healthcare Risk Management*, August, 90-91.
7. Clapper TC, Kong M(2012) TeamSTEPPS: The patient safety tool that needs to be implemented. *Clinical Simulation in Nursing* 8:e3670-e373.
8. Kardong-Edgren S, Adamson KA, Fitzgerald C (2010) A review of currently published evaluation instruments for human patient simulation. *Clinical Simulation in Nursing*, 6: e25-e35
9. Howard V (2012, December) Simulation evaluation. Webinar cosponsored by the American Association of Colleges of Nursing and the International Nursing Association for Clinical Simulation and Learning.
10. Almeida S (2009) Designing evaluation systems for TeamSTEPPS. Webinar sponsored by the American Institute for Research Designing Evaluation Systems for TeamSTEPPS.
11. Haertel EH (2006) Reliability in Brennan, R.L. (Ed.). *Educational Measurement* (4th ed.). National Council on Measurement in Education, American Council on Education. Westport, CT: Praeger Publishers.
12. Field A (2009) *Discovering statistics using SPSS*. Thousand Oaks, CA: Sage.