Contextualization and validation of the interprofessional collaborator assessment rubric (ICAR) through simulation: Pilot investigation

Fatemeh Keshmiri¹, Sari Ponzer², AmirAli Sohrabpour³, Shervin Farahmand⁴
Farhad Shahi⁵, Shahram Bagheri-Hariri⁶, Kamran Soltani-Arabshahi⁷
Mandana Shirazi*⁸, Italo Masiello⁹

Received: 30 September 2015 Accepted: 27 January 2016 Published: 1 August 2016

Abstract

Background: Simulation can be used for educating, evaluating and assessing psychometric properties of an instrument. The aim of this study was to contextualize and assess the validity and reliability of the Interprofessional Collaborative Assessment tool (ICAR) in an Iranian context using simulation.

Methods: In this descriptive study, contextualization of the ICAR was assessed through several steps. Firstly, validity assessment was approved through expert panels and Delphi rounds. Secondly, reliability assessment was done by arranging a simulation video and assessing reproducibility, test-retest (ICC), internal consistency (Cronbach’s Alpha) and inter-rater reliability (Kappa). The participants included 26 experts, 27 students and 6 staff of the Standardized Simulation Office of Tehran University of Medical Sciences.

Results: Contextualization and validity of the ICAR were approved in an Iranian context. The reliability of the tool was computed to be 0.71 according to Cronbach’s Alpha. The test-retest was calculated to be 0.76.

Conclusion: The Iranian ICAR can be a useful tool for evaluating interprofessional collaborative competencies. The development of the instrument through a simulation scenario has been a positive prospect for researchers.

Keywords: Interprofessional Collaboration, Interprofessional Education, Assessment, Validation.


Introduction

Learning to work interprofessionally is fundamental to patient safety and a patient-centered approach (1). According to the World Health Organization’s definition, interprofessional education (IPE) is the process by which a group of students or health care providers of different professions learn with, from and about each other (1).

In IPE programs, a competency-based approach is often used to assess the achievements and quality of interprofessional work (2-3). The components of a
Validation of ICAR

competency-based assessment instrument should include objectives of assessment and its guidelines (4). Some instruments, designed to assess interprofessional competences, have been developed and validated (5-8). In his study, Oates states that Interprofessional Collaborator Assessment Rubric (ICAR) can be used to assess knowledge/skills and behavioral acquisition (8). The aim of this study was to validate the ICAR (6,9-10) as an interprofessional performance assessment tool in an Iranian context.

The properties of the ICAR can be favorable to examiners, instructors and learners because they can increase the quality of instruction, and their functionality as a guideline with respect to the learning expectations and fostering critical reflection on performance (6). However, to assess interprofessional competencies, the researcher has to assess teamwork skills and communication in interaction with other professional members. This remains to be difficult since those assessed have to show effective collaboration and information exchange (11,12). Hence, several issues can affect the assessment, validity and reliability of a tool. Therefore, the quality assurance of a tool should be meticulously run and adapted to different contexts.

The validation process is usually adapted in a real educational context (6), and it seldom is applied in a simulation setting (13-17). Considering the challenges of assessing the reliability of interprofessional collaboration (10,13-14), the aim of this study was to contextualize and assess the validity of the ICAR in an Iranian context, and estimate the reliability of the ICAR reproducibility, internal consistency and inter-rater reliability (6) through simulation.

Methods

In this descriptive study, the ICAR was contextualized and validated from December 2011 to January 2013 (Fig. 1). The original instrument, which is based on a 4-point Likert scale, includes 31 items within six domains: Communication, Collaboration, Roles and Responsibilities, Collaborative Patient/Client-Family Centered Approach, Team Functioning and Conflict Resolution/Management domains. Scores indicate how often specific interprofessional competencies occur; i.e., score of 1 indicates minimally, score of 2 indicates developing (occasionally), score of 3 designates competent (frequently) and score of 4 represents mastery (consistently) (6).

Study Area

This study was conducted at Tehran University of Medical Sciences in Iran, a developing country in the Middle East. In Iran, medical education systems are centralized and medical curriculums are developed by Ministry of Health. However, interprofessional education programs are not considered in the curriculum. Based on the authors’ Knowledge, interprofessional education and collaboration are not considered in the curriculum of Middle Eastern countries. In Iran, running IPE is not an easy task due to the cultural boundaries to the doctor-center discipline. Running IPE in Iran needs some preparation of context; therefore, this study, which contextualized tools and developed educational materials, may be used as a foundation for establishing new IPE system in Iran. This study was conducted to prepare the infrastructure of education and evaluate interprofessional education in Iran.

Participants

The total number of participants in this study during the contextualization and validation phases were 59; of whom, 26 were experts, 27 students, two simulation educators, two standardized patients and two were film producers. In this study, experts were individuals who had at least 10 years of experience in clinical education, non-technical skills training (such as communication and teamwork) and were familiar with IPE.

The process is described in Figure 1.
**Study Design**

**Contextualization of the instrument:** The process was based on the “Toolkit on Translating and Adapting Instruments” (18). The instrument was adapted to the Iranian context through the following process:

1. Translating from English to Farsi
2. The tool was reviewed and adjusted by the expert panel.
3. The tool was back-translated separately from Farsi to English by translators and approved by the committee members.

**Validity Assessment of the ICAR, Part 1: English to Farsi:** The expert panel confirmed the external validity of the tool following the translation from English to Farsi and through applying the guidelines of the Human Sciences Research Institute (HSRI) (18). Content and face validity were as-
simulated scenario was recorded (36 hours).

C-parametric analysis tests (Kappa coefficient (response rate was 80%).

The simulation educators were involved in the scenario to increase the u-play. Simulation educators were involved in the scenario to increase the u-play.

The agreement between the first and second Delphi round. The agreement between the first and second Delphi round. The agreement between the first and second Delphi round. The agreement between the first and second Delphi round. The agreement between the first and second Delphi round. The agreement between the first and second Delphi round. The agreement between the first and second Delphi round. The agreement between the first and second Delphi round. The agreement between the first and second Delphi round. The agreement between the first and second Delphi round. The agreement between the first and second Delphi round. The agreement between the first and second Delphi round. The agreement between the first and second Delphi round. The agreement between the first and second Delphi round. The agreement between the first and second Delphi round. The agreement between the first and second Delphi round. The agreement between the first and second Delphi round. The agreement between the first and second Delphi round. The agreement between the first and second Delphi round. The agreement between the first and second Delphi round. The agreement between the first and second Delphi round. The agreement between the first and second Delphi round. The agreement between the first and second Delphi round. The agreement between the first and second Delphi round. The agreement between the first and second Delphi round. The agreement between the first and second Delphi round.

The reliability of the video and authenticity of the simulation were approved in terms of interprofessional collaboration skills.

Reliability Assessment of the ICAR: In this study, the reliability of the tool was assessed by performing reproducibility (test-retest), internal consistency and inter- rater reliability analyses. To do so, two rater-training sessions were provided. In the first session, raters gained familiarity with the reliability assessment process, the ICAR and its guideline. Then, the raters watched the video produced in the earlier phase and completed the Iranians’ ICAR. Afterwards, they received feedback from the main investigator of the research (M SH).

The test-retest approach was done by the raters filling out the ICAR with one-week interval between the first and second occasions (20). Internal consistency was calculated per domain through Cronbach’s Alpha. The agreement between the 23 raters and experts’ score was assessed through correlation (21).

Inter-rater reliability was approved by assessing the correlation between four expert raters after watching the video and applying non-parametric analysis tests (Kappa coefficient) (22,23).

External Validity Assessment, Part 2: Farsi to English: After demonstrating the validity and reliability of the ICAR in an Iranian context, the tool was back translated separately from Farsi to English. The translated version was sent to the main developer, Professor Curran, who confirmed the consistency of the final ICAR. The reproducibility of the ICAR was assessed by ICC analysis, the most appropriate and commonly utilized test for assessing parametric data (21,24,25). In this study, the test-retest approach was measured by the ICC, internal consistency was measured by Cronbach’s Alpha and inter-rater reliability was assessed by Kappa coefficient. The agreement between the raters and experts’ scores was assessed using descriptive statistical analysis. SPSS Version 16 was used.

The reliability of the tool was assessed by performing reproducibility (test-retest), internal consistency and inter-rater reliability analyses. To do so, two rater-training sessions were provided. In the first session, raters gained familiarity with the reliability assessment process, the ICAR and its guideline. Then, the raters watched the video produced in the earlier phase and completed the Iranians’ ICAR. Afterwards, they received feedback from the main investigator of the research (M SH).

The test-retest approach was done by the raters filling out the ICAR with one-week interval between the first and second occasions (20). Internal consistency was calculated per domain through Cronbach’s Alpha. The agreement between the 23 raters and experts’ score was assessed through correlation (21).

Inter-rater reliability was approved by assessing the correlation between four expert raters after watching the video and applying non-parametric analysis tests (Kappa coefficient) (22,23).

External Validity Assessment, Part 2: Farsi to English: After demonstrating the validity and reliability of the ICAR in an Iranian context, the tool was back translated separately from Farsi to English. The translated version was sent to the main developer, Professor Curran, who confirmed the consistency of the final ICAR. The reproducibility of the ICAR was assessed by ICC analysis, the most appropriate and commonly utilized test for assessing parametric data (21,24,25). In this study, the test-retest approach was measured by the ICC, internal consistency was measured by Cronbach’s Alpha and inter-rater reliability was assessed by Kappa coefficient. The agreement between the raters and experts’ scores was assessed using descriptive statistical analysis. SPSS Version 16 was used.
Ethical Considerations
The Ethical Committee of Teheran University of Medical Sciences (ID: 91-01-30-17052) approved this study. Prof. Curran gave permission to use the ICAR instrument. Informed consent was taken for all the raters who participated in this study.

Results
Demographic Characteristics of the Study Participants
The mean±SD age of the expert panel members was 43.5±6.5yrs., and their mean±SD work experience was 10±8.5yrs. In this study, 50% (n=6) of the participants of Delphi rounds were men. The participants' mean±SD age was 41.5±8.5 yrs., and their mean±SD of work experience was 12±2.5yrs. The mean (Mean±(SD)) age of the video production committee was 37.5±9.6yrs., and their mean±SD of work experience was 10±8.5yrs. The mean±SD age of the four raters in inter-rater reliability assessment was 33.2±6.9yrs. The majority of the raters in reproducibility assessment were female (52%) and their mean±SD age was 28.4±2yrs.

Validation of the ICAR
The content and face validity of the Iranian ICAR was approved through two Delphi rounds. The validation of scenario and video production was approved through consensus.

Psychometric Properties
The internal consistency according to Cronbach’s Alpha was 0.71, and the test–retest reliability was 0.76 (Table 1). The highest test-retest reliability was in the team function domain (ICC=0.83), and the lowest in the domain of communication (ICC=0.73). The agreement between the raters and the experts’ score was calculated to be 67.8 and 84.3(Table 1). The inter-rater reliability was calculated by kappa coefficient and turned out to be K=0.7 (Table 2).

Discussion
Measuring interprofessional collaboration

<table>
<thead>
<tr>
<th>ICAR domains</th>
<th>Test- Retest: ICC</th>
<th>Internal Consistency: Cronbach’s Alpha</th>
<th>Agreement between raters’ scores (23) and experts’ score (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Communication</td>
<td>0.73</td>
<td>0.65</td>
<td>75.12</td>
</tr>
<tr>
<td>Roles and responsibilities</td>
<td>0.80</td>
<td>0.71</td>
<td>72.65</td>
</tr>
<tr>
<td>Collaboration</td>
<td>0.75</td>
<td>0.74</td>
<td>84.34</td>
</tr>
<tr>
<td>Collaborative patient centered approach</td>
<td>0.75</td>
<td>0.71</td>
<td>77.17</td>
</tr>
<tr>
<td>Team function</td>
<td>0.83</td>
<td>0.75</td>
<td>74.78</td>
</tr>
<tr>
<td>Conflict management</td>
<td>0.75</td>
<td>0.75</td>
<td>67.80</td>
</tr>
<tr>
<td>ICAR Domains</td>
<td>0.76</td>
<td>0.71</td>
<td>75.31</td>
</tr>
</tbody>
</table>
competencies is still a challenge; therefore, applying a valid and reliable instrument remains a significant issue. The findings of this study confirmed the innovative validation process by endorsing the reliability of the ICAR within a simulated setting. In this study, several professionals were engaged in the research to be in line with the assumption of multidisciplinary attribution of IPE.

In our study, the tools' validity was strongly agreed upon by applying the accepted HSRI's guideline (18), Delphi rounds, an expert panel, and receiving feedback from the instrument’s main author (6), which was in line with the research procedure of other studies (26,27). We followed the research process by Curran et al. (6) and applied two rounds to compile the instrument in a rigorous fashion just as other researchers have done (28).

We used simulation to assess the psychometric properties (reliability) of the instrument in the interprofessional education domain. The controlled setting allowed the researchers to assess the interprofessional skills by controlling for confounding factors and possible Hawthorne effect that may occur in a real setting. In their study, Hayward et al. (10) assessed the reliability of another interprofessional performance tool in a real context, but despite the extended observation, they were not able to assess all the domains of their instrument. In this study, to avoid the challenges, reproducibility was assessed within a simulated situation. Shirazi et al. (29) conducted a research to evaluate the reliability of standardized patients (SP) and to assess communication skills as opposed to standard raters. Despite the limited number of assessors, which was a limitation, the researchers found an acceptable correlation among all raters (29). In this study, to increase the statistical power and obviate potential biases, the number of raters was increased to 23.

The highest internal consistency was in the domain of team function and challenge management. The lowest internal consistency was computed in the communication domain. This could be interpreted as an inconsistency between items because of different issues such as ‘respect’, which are more subjective and more dependent on raters’ perceptions and experience. However, in accordance with this latest finding, studies from Shyne et al. (30) and Najafi et al. (31) found similar low internal consistency with the Team STEPPS Teamwork Attitudes Questionnaire (T-TAQ) and SDOT, respectively. In contrast with our findings, Hayward et al. study (10) found internal consistency of communication to have the highest domain. This may be due to its use of a multi-source feedback (MSF) process to assess collaboration competencies (10).

The highest agreement between the 23 raters and the experts’ scores was in the field of collaboration (84.3%), which could be related to the raters’ familiarity with these concepts and gaining experience in the group work. Moreover, the items’ construct in this domain consisted of simplicity, clarity and relevance of items for the raters. The lowest accordance was in the domain of conflict management (67.8%). These findings may be due to the possible limited involvement of the 4th-year medical students in health care management issues or lack of training. However, this information could not be confirmed.

The correlation between experts’ rating was approved by non-parametric statistical tests. The Kappa coefficient (32) demonstrated propriety agreement among each pair raters and between all raters’ (four) scores. In line with our study, findings from Schmitz demonstrated acceptable or highly acceptable coefficients (33). However, in contrast to our findings, Boulet et al. found a low correlation among their raters (0.09 and 0.29), which can be due to the lack of training or an assessment guide in their study (34). In this study, to inhibit the possible low inter-rater reliability among the raters, the research group compiled a guideline for reliability assessment and conducted training sessions.
Limitations

The validity assessment of an interprofessional assessment tool in a real setting is a difficult task as there are many confounders. Moreover, in Iran, we do not have interprofessional wards where we could run such studies. Therefore, in this study, we used a video of a simulated setting to assess the ICAR psychometric properties. Although this was a psychometric study, one of the limitations was the reproducibility assessment of the instrument by medical students due to the restriction in the implementation of the interprofessional meetings. Limited generalizability across institutions, countries and even within the setting of an institution (i.e., beyond the trauma team) might be a limitation.

Conclusion

Assessing IPE is a complex task due to the difficult task of assessing team competencies. This study confirmed the validity and reliability of the modified ICAR to evaluate interprofessional collaborative competencies in an Iranian context, which was done using a novel process of using a simulated scenario. IPE in an Iranian context is a new endeavor and we recommend conducting more research on this topic.

Acknowledgements

Sincere thanks go to Professor Curran for his comments and kind contribution and also to the medical students at TUMS who took part in this research.

References


