



## Original Article

# The Reliability and Validity of the Korean Version of the Indiana University Simulation Integration Rubric for Interprofessional Communication among Nursing and Medical Students

Chae, Shinhye<sup>1)</sup> · Kim, Jae-Bum<sup>2)</sup> · Shon, Soonyoung<sup>3)</sup>

- 1) Graduate Student, College of Nursing, Keimyung University, Daegu, Korea  
2) Professor, College of Medicine, Keimyung University, Daegu, Korea  
3) Assistant Professor, College of Nursing, Keimyung University, Daegu, Korea

**Purpose:** This study aimed to translate and validate the Korean version of the Indiana University Simulation Integration Rubric (K-IUSIR) for assessing interprofessional communication in simulation-based education, focusing on its reliability and validity. **Background:** Interprofessional education (IPE) improves communication and teamwork skills critical for ensuring patient safety. However, there is a lack of standardized tools for assessing interprofessional communication in Korea. **Methods:** A methodological study involving a secondary analysis was conducted with 221 nursing and medical students. Data were collected through simulation recordings and surveys and analyzed using Cronbach's  $\alpha$ , intraclass correlation coefficients (ICCs), expert content validity, and Rasch analysis. **Results:** The K-IUSIR demonstrated high internal consistency (Cronbach's  $\alpha=0.854$ ), inter-rater reliability (ICC=0.832), and strong content validity. Rasch analysis confirmed appropriate item fit and difficulty levels, with a 3-point Likert scale identified as the most suitable format. **Conclusion:** The K-IUSIR is a reliable and valid tool for evaluating interprofessional communication, which can be utilized in developing interprofessional education programs. Furthermore, it can facilitate better interprofessional collaboration through effective communication.

**Key Words:** Communication; Interprofessional education; Medical student; Nursing student; Validation study

\*This study was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (#RS-2023-00210104).

Received Dec 10, 2024 Revised Feb 7, 2025 Accepted Feb 16, 2025  
Corresponding author: Shon, Soonyoung <https://orcid.org/0000-0003-0571-2181>  
College of Nursing, Keimyung University  
1095 Dalgubeol-daero, Dalseo-gu, Daegu 42601, Korea  
Tel: +82-53-258-7657, Fax: +82-53-258-7616, E-mail: sy.shon@kmu.ac.kr

## INTRODUCTION

The complexity of modern healthcare environments requires a multidisciplinary approach and effective collaboration among healthcare professionals to ensure patient safety [1]. Approximately 30% of medical errors are attributed to communication and teamwork issues, and effective teamwork and communication are essential for patient safety [2]. The World Health Organization (WHO) emphasizes interprofessional communication as a critical component of patient safety. In response to the growing demand for interprofessional collaboration, theoretical foundations for interprofessional education (IPE) have been developed and integrated into healthcare training programs [3]. IPE is a key educational strategy designed to enhance interprofessional collaboration among healthcare professionals. It improves teamwork, communication, and problem-solving skills while fostering knowledge, competencies, and attitudes necessary for effective collaboration [2,4]. Additionally, IPE fosters positive mutual perceptions and attitudes among different professional groups, contributing to higher quality care and supporting the delivery of patient-centered, coordinated healthcare [5].

The Interprofessional Education Collaborative (IPEC) defines the core competencies of healthcare professionals to be enhanced through interprofessional education as ethics, values, roles and responsibilities, communication, and teamwork [6]. These core competencies serve as benchmarks for evaluating the effectiveness of interprofessional education. Simulation-based Interprofessional Education (Sim-IPE) has gained attention as an effective method for enhancing these competencies [7]. Sim-IPE improves interprofessional attitudes, communication skills, teamwork, and readiness for interprofessional education. Furthermore, it offers participants opportunities to collaborate across professions in simulated settings, fostering an understanding of each profession's role, and enabling the practical application of interprofessional collaboration [8].

Therefore, an appropriate tool to evaluate the effectiveness of IPE is critical. Evaluations typically focus on the core competencies of IPE and use frameworks, such as Kirkpatrick's evaluation model, which categorizes outcomes into four levels: reaction, learning, behavior, and results [9]. The Indiana University Simulation Integration Rubric (IUSIR), developed by Reising et al. [10], is suitable for assessing both learning and behavior. Learning evaluations assess the extent to which learners achieve their intended objectives and the changes that result from their participation in educational activities. Behavioral evaluations measure the application of acquired knowledge and

skills in professional practice [9].

The IUSIR consists of two domains-“individual” and “team”-encompassing a total of 12 items. This tool demonstrated high internal consistency and validity during the developmental phase. However, the IUSIR has been criticized for its narrow focus, as it assesses only the interprofessional communication domain and does not evaluate broader interprofessional competencies [11]. Additionally, its applicability in real-world clinical settings beyond simulation scenarios has yet to be confirmed [11]. Despite these limitations, the IUSIR is widely recognized for its high reliability in various studies and praised for its ability to differentiate performance based on individual abilities [11]. Furthermore, as an observer-based evaluation tool, it offers greater objectivity than self-reported measures and is considered a practical instrument for simultaneously assessing both team and individual performance.

In Korean healthcare education, IPE is regarded as a “key element transforming 21st-century healthcare education” [12]. However, the implementation and research of IPE in Korea remain in the early stages. Standardized tools to evaluate the effectiveness of interprofessional education in Korea are lacking [12]. The tools currently used are primarily adaptations or translations of those developed in other countries, which present limitations due to cultural and linguistic differences. Communication is a complex process that extends beyond simple verbal exchanges, involving cultural norms, values, and interaction patterns. Tools developed in English are likely to reflect the communication styles and cultural characteristics of English-speaking contexts [13]. Applying these tools directly to Korean users may fail to capture the important nuances and communication styles specific to Korean culture. Therefore, adapting and validating communication assessment tools in Korean is crucial for improving the accuracy and reliability of evaluations and ensuring that they reflect Korean speakers' cultural and communicative characteristics.

Currently, validated tools for evaluating the effectiveness of IPE in South Korea are limited to measuring attitudes toward interprofessional collaboration [12] and self-efficacy in interprofessional learning [14]. These tools assess learners' attitudes and perceptions toward educational programs. However, reaction evaluation alone is insufficient to fully capture what learners have gained from the program (learning), how they apply it in practice (behavior), or the program's impact on organizational or patient outcomes (results). While evaluating attitudes and self-efficacy provides a valuable starting point, a comprehensive

evaluation of IPE effectiveness requires a multi-level approach encompassing learning, behavior, and results.

The IUSIR facilitates this multi-level approach by thoroughly assessing both individual and team competencies. It offers a practical framework for analyzing IPE outcomes and identifying areas for improvement. Additionally, the IUSIR bridges the gap between simulation-based education and real-world clinical practice by fostering the effective application of interprofessional skills in professional settings. This study aimed to determine whether the IUSIR, a tool gaining increasing attention for IPE, can be applied reliably and validly in Korea as an effective evaluation instrument, contributing to a robust IPE assessment framework.

## METHODS

### 1. Design

This study employed a methodological design with a secondary analysis to validate the reliability and validity of the Korean version of the IUSIR tool. The validation process involved analyzing recorded simulation videos from an interprofessional education program.

### 2. Participants

The participants in this study were fourth-year nursing and medical students who voluntarily participated in interprofessional education programs at K University between 2022 and 2023. Recorded videos of participants who provided consent for data use, as documented through signed consent forms and data utilization agreements obtained during the program, were analyzed. Based on the sample size criteria, 221 participants were included in the study, following the guidelines that 200 participants are suitable for Rasch analysis when the number of items is 40 or fewer [15] and that five to 10 times the number of items is considered appropriate [16]. Simulations were conducted in teams of three or four members, resulting in 56 recorded videos and 221 individual survey responses included in the analysis. The simulation was conducted to assess responses to acute exacerbation emergencies in adult patients. It lasted approximately 10 to 15 minutes and included a balanced ratio of medical and nursing students.

### 3. Measurements

In this study, we collected data on the general characteristics of participants in the educational program as well as

their scores on the Indiana University Simulation Integration Rubric (IUSIR). The personal characteristics assessed included age, major, sex, and prior experience with Interprofessional Education (IPE). The IUSIR is specifically designed to measure communication among professionals during simulation exercises and was developed through research conducted with nursing and medical students. It comprises two domains: "Individual" and "Team," with a total of 12 items. Each item is rated on a 3-point scale, where "Below Average" receives 1 point, "Average" receives 3 points, and "Above Average" receives 5 points; intermediate scores can also be assigned as needed.

The tool evaluates communication skills based on scores assigned by the evaluators, with higher scores indicating better communication abilities. In the individual domain, the evaluation items included nonverbal communication, communication skills, feedback integration, expressing opinions during decision-making, identifying patient issues, and communicating with patients. The team domain assesses factors, such as team energy and organization, intra-team communication skills, integration of team members' opinions, clinical situational awareness, patient education, and patient reassessment.

The IUSIR demonstrated reliability during its development, with a reported Cronbach's  $\alpha$  ranging from 0.79 to 0.90. Positive correlations were found among the items and between individual items and overall scores, confirming the tool's reliability. For validity, the IUSIR was tested to determine whether it could detect differences in performance levels between students by comparing junior and senior undergraduate students of health professionals. The results indicated that senior students achieved statistically significantly higher scores than junior students, thus demonstrating its validity.

### 4. IUSIR Translation and Content Validity

The study was conducted with approval from the original author to use the IUSIR tool. The translation process followed WHO-recommended guidelines [17]. First, the original English tool was translated into Korean. Subsequently, a bilingual translator fluent in both English and Korean, but unfamiliar with the original tool, back-translated the Korean version into English. Two nursing professors and one medical professor specializing in medical education reviewed the translated tool. They evaluated the equivalence of meaning, consistency, accuracy, and contextual relevance of the translation. Finally, the original tool and the back-translated content were compared to align with the developer of the original tool.

Based on feedback from the expert review, the items were revised and supplemented to better fit the context of our country, resulting in the completion of the preliminary translated version. A four-point Likert scale, ranging from “not suitable at all” to “very suitable,” was used to evaluate the items. The Content Validity Index (CVI) was calculated based on the proportion of items rated as “suitable” or “very suitable” by the experts.

A preliminary survey was conducted using five IPE simulation videos to evaluate inter-rater reliability. IPE was implemented for fourth-year nursing and medical students. The simulation scenarios focused on managing acute deterioration and emergencies in adult patients. The training aimed to enhance interprofessional teamwork, effective communication during shock situations, and prompt and appropriate responses. Multidisciplinary teams of nursing and medical students participated in simulations designed to guide decision-making and actions based on the patient’s condition. Additionally, the simulation provided opportunities for hands-on practice with medications and medical equipment.

The raters, consisting of two nurses with master’s degrees and one nursing professor (three individuals, including the researcher), received prior training on the items and criteria of the K-IUSIR tool. Evaluations were conducted simultaneously under identical video viewing conditions. After three detailed debriefing sessions to review the consistency of scores and interpretations, five IPE simulation videos were assessed to ensure inter-rater reliability. Through this process, the final version of the Korean Indiana University Simulation Integration Rubric (K-IUSIR) was completed.

## 5. Ethical Consideration

This study was approved by the Institutional Review Board (IRB No. DSMC 2024-04-033). This secondary analysis exclusively utilized data from participants who provided prior consent. Computerized data were secured with password protection, limiting access to researchers only.

## 6. Data Analysis

Data were analyzed using SPSS version 29.0 (IBM, Armonk, New York, USA) and Winsteps version 5.8.2 (Winsteps Inc., Chicago, IL, USA). The general characteristics of the study participants were analyzed using frequencies and percentages. Per Classical Test Theory (CTT), the reliability of the measurement instruments was assessed using Cronbach’s  $\alpha$ , whereas inter-rater reliability

was analyzed using the intraclass correlation coefficient (ICC). The validity of the measurement instruments was examined through the Content Validity Index, utilizing a panel of experts. Construct validity was evaluated through Rasch analysis based on Item Response Theory (IRT), assessing item fit, item difficulty, and the appropriateness of the number of response categories.

IRT is a statistical method used to evaluate validity and reliability by examining the relationship between item characteristics and respondent abilities [18]. The Rasch model improves on CTT by transforming ordinal into interval scales, allowing for more accurate score interpretation. CTT estimates item difficulty and respondent ability dependently; meanwhile, the Rasch model analyzes them independently, enhancing reliability, and also evaluates response category functionality, helping to refine survey tools [19]. As a result, the Rasch model offers a more objective and reliable alternative to CTT for data analysis [19].

A fundamental assumption of Rasch analysis is that the measurement tool is unidimensional, meaning that it measures a specific trait. Unidimensionality is assessed through principal component analysis of standardized residuals, where an explained variance of more than 30% [19], and an eigenvalue of less than 3.0 for the first or second residual variance (excluding the Rasch factor) supports the unidimensionality of the test items [20]. Additionally, in terms of item polarity, a point-measure correlation coefficient of .3 or higher was considered supportive of unidimensionality.

Item fit evaluates how well the item response data align with the expected values of the Rasch model, verifying whether a specific item functions appropriately for the measurement objectives [18]. The primary criteria for assessing item fit were infit and outfit mean square fit statistics (MNSQ) [21]. Generally, an MNSQ value close to one indicates a suitable item. According to observational study criteria, the acceptable fit ranges from 0.5 to 1.7; items with a fit index greater than 1.7 are deemed inappropriate because of poor alignment with the factors, while items with a fit index less than 0.5, indicate repetition of poorly fitting items with minimal variation [22]. The point-measure correlation coefficient also reflects item fit and should be above .3 to support unidimensionality [20]. Separation reliability assesses the consistency with which difficulty levels are distinguished, with values ranging from 0 to 1, and higher values indicating greater distinguishability [19]. The separation index measures the discrimination ability of the evaluation targets, with higher values indicating finer differentiation [19]. Generally,

separation reliability of .8 or higher, combined with a separation index of 2.0 or greater, is considered acceptable [20].

Item difficulty measures typically range from -3 to +3, with values closer to 0 indicating average difficulty [23]. According to established criteria for evaluating item difficulty, an index of -0.5 or lower is interpreted as low-difficulty, while an index of +0.5 or higher indicates high-difficulty items [24]. Item difficulty can be assessed using an item-person map, where items positioned higher in the distribution are considered more difficult and those positioned lower are considered easier [25]. The appropriateness of the number of response categories must meet several criteria. Each category should have an observed count greater than 10, and the observed average measurement should progressively increase with each category. Additionally, each category's outfit MNSQ score must be less than 2.0. The step calibration index should also increase progressively with each category, and the absolute difference in the increasing step adjustment index should be between 1.40 and 5.00 for the categories to be considered appropriate [19]. The suitability of the response category scale was visually evaluated using probability curves. Consistent intersection points among the scales indicated favorable conditions and clearly distinct areas for each category [21].

## RESULTS

### 1. General Characteristics

The study participants consisted of 82 males (37.1%) and 139 females (62.9%), with ages ranging from 22 to 33 years, and an average age of 25.27 years. Among the participants, 116 (52.5%) were nursing students and 105 (47.5%) were medical students. Ten students (4.5%) had prior experience with interprofessional education, and 165 students (74.7%) had attended courses in other departments. In addition, 64 students (29.0%) reported participating in joint club activities (Table 1).

### 2. Reliability of IUSIR

#### 1) Internal consistency

Cronbach's  $\alpha$  for the IUSIR was .85, indicating good internal consistency.

#### 2) Inter-rater reliability

The three raters familiarized themselves with the evaluation criteria and methods through a preliminary study

**Table 1.** General Characteristics of Participants (N=221)

Characteristics	Categories	n (%) or M±SD
Gender	Men	82 (37.1)
	Women	139 (62.9)
Age (year)		25.27±1.96
Major	Nursing	116 (52.5)
	Medicine	105 (47.5)
Interprofessional education experience	Yes	10 (4.5)
	No	211 (95.5)
Experience of taking courses in other departments	Yes	165 (74.7)
	No	56 (25.3)
Experience of interdisciplinary club activities	Yes	64 (29.0)
	No	157 (71.0)

and independently analyzed 56 IPE simulation videos, dividing them into three equal parts. The results of the inter-rater reliability assessment among the three raters showed an ICC of .832, indicating a high level of reliability among the raters.

### 3. Validity of IUSIR

#### 1) Content validity

The content validity was established through the participation of six experts. Three of the 12 items had an item-level content validity index (I-CVI) of .833, while the remaining items had an I-CVI of 1.0. The scale-level content validity index (S-CVI) was .955, indicating strong content validity. Based on the expert panel's feedback, the items' wording was revised for clarity and consistency. For instance, "document materials" was changed to "presented materials," and "search for opinions" was revised to "seek opinions." Additionally, specifics regarding team interaction positivity and ambiguity in describing clinical situations were enhanced, and redundancies and item appropriateness were reevaluated. This process improved both translation accuracy and the overall quality of the items.

#### 2) Construct validity

Construct validity was assessed using Rasch analysis based on the IRT. Initially, the fundamental assumption of unidimensionality was verified to determine whether the Rasch analysis could be applied. In the principal component analysis of the residuals conducted for this assessment, the variance explained by Rasch measurement was 42.7%. The eigenvalues for the first and second residuals,



**Table 2.** Results of Item Difficulty

(N=221)

Item No.	Difficulty	Infit MNSQ	Outfit MNSQ	Point measure correlation
6	0.36	1.73	1.69	0.59
11	0.05	1.45	1.31	0.60
8	0.84	1.15	1.15	0.50
10	-0.18	1.15	1.13	0.58
1	-0.64	0.85	0.91	0.55
7	-0.34	0.90	0.85	0.56
2	0.54	0.89	0.89	0.62
3	-0.18	0.80	0.85	0.65
5	0.17	0.73	0.82	0.69
12	-0.72	0.82	0.76	0.62
9	0.08	0.73	0.79	0.59
4	0.03	0.64	0.65	0.73

Person reliability: .80  
Item reliability: .95

Person separation index: 2.00  
Item separation index: 4.52

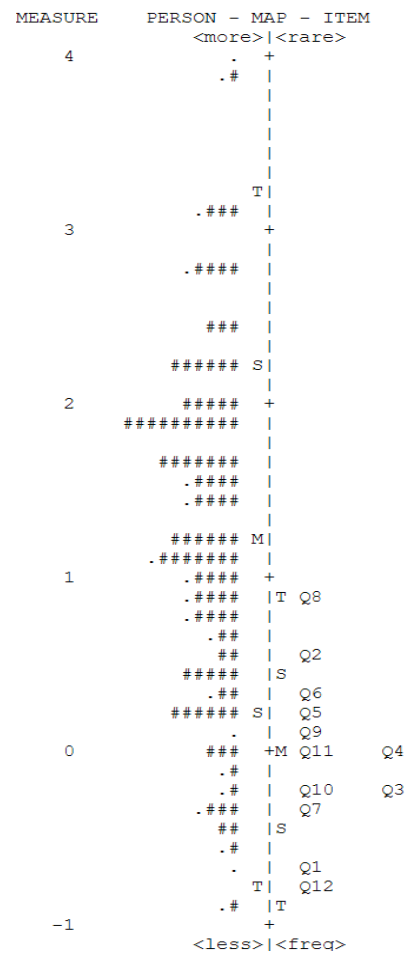
MNSQ=mean square fit statistic.

excluding the Rasch factor, were 2.67 and 1.77, respectively, both below 3.0, confirming unidimensionality. Additionally, the point-measure correlation coefficients for the 12 items ranged from .50 to .73, further supporting unidimensionality.

Table 2 presents item fit and item difficulty. The analysis revealed that the infit and outfit MNSQ values ranged from 0.62 to 1.73, indicating that Item 6 was considered unsuitable. The personseparation reliability was .80, and the separation index was 2.00, whereas the item separation reliability was .95, with an item separation index of 4.52. Overall, these results indicated that the separation reliability was excellent.

According to the criteria for evaluating item difficulty [24], two items (Items 1 and 12) were classified as low difficulty, with indices below -0.5. Eight items (Items 3, 4, 5, 6, 7, 9, 10, and 11) fell into the moderate difficulty category, with indices ranging from -0.5 to +0.5, whereas two items (Items 2 and 8) were classified as high difficulty, with indices of +0.5, or higher. This distribution indicated an appropriate range for item difficulty (Table 2). Item 8, which pertained to communication skills within a team, had the highest difficulty, whereas item 12 was related to patient reassessment. The distribution of items and respondent measurement scores showed that most participants clustered within the moderate difficulty range (1~2 logit), and there was an overlap between the distribution of respondents and items (Figure 1).

This tool was developed using a three-point scale but allowed for the assignment of a middle score as needed.

**Figure 1.** Item-person map.

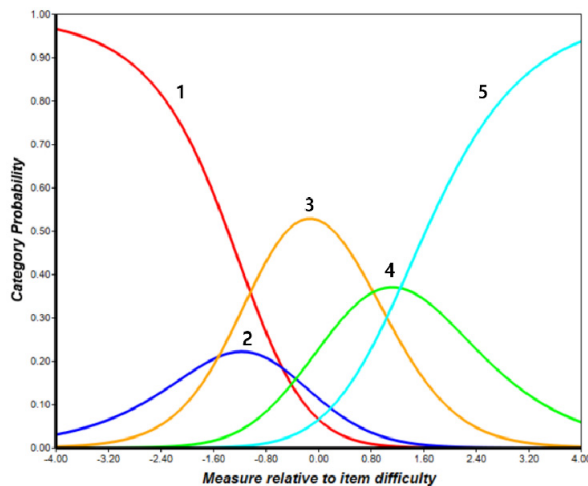
Therefore, a moderate score was included in the evaluation. However, analysis of the five-point scale revealed that step calibration did not increase progressively, with some absolute differences in the step adjustment values below 1.4, indicating that the five-point scale was unsuitable (Table 3). As a result, the tool was reanalyzed using a three-point scale, which confirmed the appropriateness of the response categories (Table 3). Furthermore, analysis of the item response category curves showed that the threshold parameters within the five-point scale category curves were not distinctly separated and did not maintain equal intervals (Figure 2-A). In contrast, the three-point scale exhibited clearly distinguishable threshold parameters, maintaining equal intervals and demonstrating suitability (Figure 2-B).

## DISCUSSION

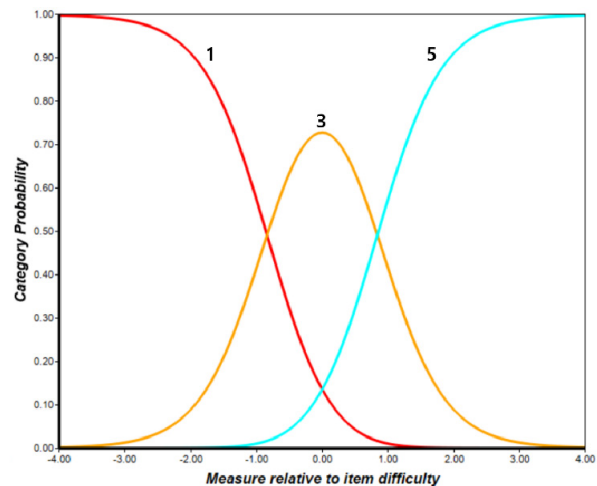
This study was a methodological and secondary analysis aimed at validating the reliability and validity of the IUSIR, a tool designed to measure communication among

experts in simulation settings. In evaluating internal consistency for reliability assessment, Cronbach's  $\alpha$  for the K-IUSIR was .853. Generally, a Cronbach's  $\alpha$  value between .80 and .90 indicates a high-reliability tool [26], suggesting that the items consistently measure communication skills among professionals. The Cronbach's  $\alpha$  of this tool is similar to the level reported in Reising's study [10], which focused on nursing and medical students, but lower than that found in Keiser's research [27], which utilized a virtual web-based platform for interprofessional education among nursing and medical students.

Inter-rater reliability was assessed using the ICC for the K-IUSIR scores measured by the three raters, yielding an ICC of .832. An ICC value above .80 indicates good reliability, suggesting that the tool maintained a high level of consistency among raters [28]. Inter-rater reliability is especially crucial in observation-based assessments because low consistency among raters can compromise the objectivity and fairness of evaluations [29]. The ICC value demonstrates that the K-IUSIR can maintain objectivity as an expert-based assessment tool by reducing assessor sub-



A. Probability curves of the five response categories.



B. Probability curves of the three response categories.

Figure 2. Probability curves.

Table 3. Scale Analysis

Categories		Observed count (%)	Average measure (logit)	Step calibration (logit)
Five-point likert	1	84 (3)	-0.26	None
	2	114 (4)	-0.11	-0.55
	3	689 (26)	0.52	-1.53
	4	733 (28)	1.26	0.83
	5	1,032 (39)	1.94	1.24
Three-point likert	1	84 (3)	0.03	None
	2	803 (30)	0.69	-1.68
	3	1,243 (67)	1.58	1.68

jectivity and establishing it as a reliable instrument. This ICC value is comparable to that reported in a study by Shin et al.[28], which analyzed observations from simulations for clinical judgment assessments. However, it is lower than the values found in interprofessional education studies conducted with nurses and physicians [2].

The content validity of the Korean-translated tool was evaluated by a panel of six experts, resulting in I-CVI values ranging from .833 to 1.0 and an S-CVI of .96. An I-CVI value of .78 or higher is considered acceptable when assessed by a panel of six to ten experts, while an S-CVI above .90 is regarded as indicating sufficient content validity [30]. It was concluded that the IUSIR maintained its content validity even after translation into Korean.

Construct validity was assessed using Rasch analysis. The variance explained by the Rasch model for unidimensionality was 42.7%, with the eigenvalues of the residual variance measured below 3.0. Additionally, the point-measure correlation coefficients were above .3 for all items. This suggests that the IUSIR is an effective tool for measuring the specific concept of communication among professionals, thereby preventing conceptual confusion among assessment items and accurately reflecting the respondents' abilities.

In the fit analysis, most items had infit MNSQ values ranging from 0.62 to 1.73, indicating their suitability for the Rasch model; however, item 6 was found to be unsuitable. These values suggest that the items functioned as intended by the model and effectively differentiated respondents based on their abilities. Generally, an MNSQ value close to 1 is considered ideal, as it indicates that the item performs consistently with the expectations of the Rasch model [21,22]. When the MNSQ value is significantly greater than 1, it suggests that the item is heterogeneous compared to other items on the scale, whereas a value below 1 indicates redundancy or similarity with other items [21,22]. Most items demonstrated a good fit with MNSQ values close to 1, supporting the tool's reliability in providing valid evaluation results even among heterogeneous populations.

Item 6, which assessed communication with patients, exhibited an MNSQ value of  $>1.7$ , indicating that it was unsuitable. The IUSIR is specifically designed to measure communication skills among healthcare professionals. Therefore, the patient communication item is conceptually misaligned with the tool's intended purpose and target population. Notably, the criteria for item fit are not absolute and can vary depending on the research objectives. For instance, some researchers may set the acceptable range for MNSQ values between 0.5 and 2.0 when assessing item

fit [23]. In this study, item 6 had an MNSQ value of 1.73, which slightly exceeded the threshold of 1.7, but remained within the acceptable range of 2.0, indicating that it was not excessively high. This allows for flexible interpretation depending on the item's research objectives and design intent.

Evaluating communication with patients provides a vital opportunity to comprehensively understand healthcare professionals' communication skills. This item complements information sharing among healthcare professionals and enhances collaborative communication within teams. Additionally, patient communication offers valuable insights into how information is exchanged within and outside of teams, effectively supporting interprofessional collaboration and patient-centered care [31]. Preliminary testing and expert validation confirmed that the inclusion of Item 6 did not compromise the reliability or validity of the K-IUSIR. Therefore, removing this item risks overlooking its potential contributions. Future studies should consider modifying this item to explicitly link patient communication with team-based communication. For instance, evaluating whether critical patient information is effectively shared with the team, conveyed at the appropriate time, and communicated with clarity can help maintain patient-centered communication while enhancing team information sharing. Implementing these refinements and reassessing Item 6 would further improve the conceptual alignment of the K-IUSIR, while also strengthening its reliability and validity.

To assess the suitability of the items, a point-measure correlation analysis was conducted, which revealed that all items of the K-IUSIR performed well. This indicates that respondents with higher characteristics scored higher on items within the factor, whereas those with lower characteristics scored lower, thereby supporting the structural validity of the scale. The separation index for respondents was 2.0, indicating that the participants could be divided into two groups. The separation reliability was .8, suggesting that the tool effectively distinguished between the respondents' abilities. The item separation index was 4.52, indicating that the items could be classified into approximately four to five difficulty groups, reflecting a diverse range of item difficulties. The item separation reliability was notably high at .95, suggesting that the items in the measurement tool effectively captured the differences in difficulty.

The item difficulty revealed that the item difficulties were evenly distributed, effectively distinguishing learners at various levels. The study participants reported relative difficulty with Items 2 and 8, which assess individual



and team communication skills, respectively. These items emphasize critical aspects of communication, a key competency in Interprofessional Education (IPE) [6], evaluating whether participants clarify their roles using clear terminology and employing closed-loop communication. It is possible that the participants were unaware of the need to clarify their roles in the simulation scenarios or did not fully understand their responsibilities. In multidisciplinary teams, roles may overlap or remain unclear, highlighting the necessity for prior training in these aspects [2].

Closed-loop communication is a structured approach designed to ensure the accuracy and clarity of information transfer. This process involves the sender delivering a message, the receiver repeating the message to confirm their understanding, and the sender verifying the receiver's interpretation [32]. Participants who did not undergo this process may have found it more challenging. Since both the senders and receivers must engage in closed-loop communication to be effective, this suggests that the entire team may have been unfamiliar with this practice. Therefore, thorough training in closed-loop communication should be provided to enhance communication skills.

The areas that participants found relatively easy were Item 1, which assessed individual nonverbal communication, and Item 12, which involved reevaluating treatment outcomes as a team. Nonverbal communication is commonly used in everyday life and tends to be more intuitive than verbal communication [33], suggesting that participants can readily demonstrate this skill without specific training. The activity of reevaluating treatment outcomes as a team is based on shared goals rather than individual subjective judgments, likely reducing the burden on team members to take sole responsibility and facilitating more active communication [34]. In contrast, the more challenging Items 2 and 8 require advanced communication skills, providing differentiation among learners, whereas the easier Items 1 and 12 are appropriate for assessing foundational communication competencies. This combination of high- and low-difficulty items underscores their contributions to evaluating learners at various levels.

The item-person map analysis indicated that respondents were primarily concentrated within the 1~2 logit range, suggesting that most exhibited moderate communication abilities. Item difficulty was largely distributed between -1 and 2 logits, effectively distinguishing respondents within this range. However, the absence of high-difficulty items in the 3~4 logit range may limit the tool's capacity to assess individuals with advanced communication skills. To address this limitation, future research should consider incorporating more challenging items.

For instance, items assessing team communication in high-stakes scenarios (e.g., clearly defining roles, collaborating efficiently, and communicating effectively during patient resuscitation) or conflict resolution skills (e.g., achieving evidence-based consensus amid professional disagreements) could be included. Integrating such items would enhance the tool's discriminative power, enabling a more precise evaluation of learners across varying competency level.

The analysis of response categories indicated that the five-point scale was unsuitable, as step calibration did not increase progressively and there was no visually even distribution among the scales. Consequently, the fit improved when adjusted to a three-point scale without a midpoint. This finding aligns with cases where reducing the scale from a previous version enhances the validity of the tool [35]. The three-point scale simplifies response options, enabling evaluators to distinguish behaviors more clearly and reducing the complexity of the assessment process. Additionally, its clear structure shortens evaluators' decision-making time, enhances the efficiency of assessments, and minimizes confusion. The well-defined boundaries between categories also facilitate a more straightforward interpretation of evaluation results and contribute to maintaining the reliability and validity of the tool. Therefore, adopting the three-point scale is recommended when using this tool.

The findings indicate that the K-IUSIR is a reliable and valid tool for assessing communication skills among professionals during simulations. It demonstrated favorable characteristics across various item properties, including difficulty, discrimination, and fit. Notably, the balance between item difficulty and discrimination allows applications across different learner levels. Additionally, the localization of the K-IUSIR into Korean has enhanced its applicability by reflecting Korea's cultural characteristics, providing a practical methodology for evaluating and applying IPE programs in diverse healthcare settings.

However, it is important to note that this study was conducted with students from a limited number of universities and within a specific simulation environment, which may have limited the generalizability of the findings. Furthermore, despite providing training to the evaluators, the possibility of evaluator bias cannot be excluded. Evaluators may have interpreted participants' communication skills based on individual perspectives or cultural nuances, which could have affected the objectivity and consistency of the scores.

Despite these limitations, the K-IUSIR remains valuable for developing communication improvement strategies,

as it systematically evaluates the core competency of IPE. Furthermore, it assesses performance (changes in learning and behavior) within the framework of Kirkpatrick's evaluation model [9]; it goes beyond the measurement of knowledge and attitudes to verify whether learning outcomes translate into actual behavior, thereby enhancing its practicality. Ultimately, the K-IUSIR serves as a critical tool for assessing improvements in communication competencies after training and can significantly contribute to the design and enhancement of interprofessional education programs. This study provides a reliable, culturally relevant tool for evaluating and improving IPE programs in Korea, contributing to advancing interprofessional collaboration education and improving healthcare services.

## CONCLUSION

A comprehensive analysis of the reliability and validity of the K-IUSIR confirms that it is suitable for assessing communication skills among professionals during simulations in South Korea. Furthermore, using a three-point scale, the K-IUSIR serves as a practical instrument, allowing for quicker and clearer evaluation of both team and individual competencies. The K-IUSIR can also be used to enhance the design and content of IPE programs by evaluating the achievement of educational goals and identifying participants' strengths and areas for improvement. Furthermore, it facilitates the early detection of communication challenges and supports the development of targeted strategies to address them, thereby strengthening participants' competencies. By promoting a culture of interprofessional collaboration, the K-IUSIR ultimately contributes to enhancing the quality of healthcare services.

The following recommendations are proposed based on the results and discussion of this study. First, as this study focused on students, future studies should revalidate the reliability and validity of the K-IUSIR with healthcare professionals using a variety of simulation scenarios. Second, to mitigate evaluator bias, evaluator training should be standardized, and evaluation guidelines should be provided to improve consistency in scoring. Lastly, further research is needed to explore the applicability of K-IUSIR beyond simulated environments to actual clinical practice situations.

## CONFLICTS OF INTEREST

The authors declared no conflict of interest.

## AUTHORSHIP

Study conception and design acquisition - Kim J-B and Shon S;

Data collection - Kim J-B and Shon S; Data analysis & Interpretation - Chae S and Shon S; Drafting & Revision of the manuscript - Chae S, Kim J-B and Shon S.

## DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Please contact the corresponding author for data availability.

## REFERENCES

1. Collaborative IE. Core competencies for interprofessional collaborative practice: 2016 update. Washington, D.C.: Interprofessional Education Collaborative; 2016.
2. Chae S, Shon S. Effectiveness of simulation-based interprofessional education on teamwork and communication skills in neonatal resuscitation. *BMC Medical Education*. 2024;24(1):602. <https://doi.org/10.1186/s12909-024-05581-1>
3. World Health Organization. Patient safety. Geneva: World Health Organization; 2003. Report No.: WHO/EIP/OSD/2003.5.
4. Guraya SY, Barr H. The effectiveness of interprofessional education in healthcare: a systematic review and meta-analysis. *The Kaohsiung Journal of Medical Sciences*. 2018;34(3):160-165. <https://doi.org/10.1016/j.kjms.2017.12.009>
5. Pascucci D, Sassano M, Nurchis MC, Cicconi M, Acampora A, Park D, et al. Impact of interprofessional collaboration on chronic disease management: findings from a systematic review of clinical trial and meta-analysis. *Health Policy*. 2021;125(2):191-202. <https://doi.org/10.1016/j.healthpol.2020.12.006>
6. Collaborative IE. IPEC core competencies for interprofessional collaborative practice: version 3. Washington, D.C.: Interprofessional Education Collaborative; 2023.
7. Kleib M, Jackman D, Duarte-Wisnesky U. Interprofessional simulation to promote teamwork and communication between nursing and respiratory therapy students: a mixed-method research study. *Nurse Education Today*. 2021;99:104816. <https://doi.org/10.1016/j.nedt.2021.104816>
8. Aldriwesh MG, Alyousif SM, Alharbi NS. Undergraduate-level teaching and learning approaches for interprofessional education in the health professions: a systematic review. *BMC Medical Education*. 2022;22:1-14. <https://doi.org/10.1186/s12909-021-03073-0>
9. Kirkpatrick D, Kirkpatrick J. Evaluating training programs: the four levels. San Francisco: Berrett-Koehler Publishers; 2006.
10. Reising DL, Carr DE, Tieman S, Feather R, Ozdogan Z. Psychometric testing of a simulation rubric for measuring interprofessional communication. *Nursing Education Perspectives*. 2015;36(5):311-316. <https://doi.org/10.5480/15-1659>

11. Brownie S, Blanchard D, Amankwaa I, Broman P, Haggie M, Logan C, et al. Tools for faculty assessment of interdisciplinary competencies of healthcare students: an integrative review. *Frontiers in Medicine*. 2023;10:1124264. <https://doi.org/10.3389/fmed.2023.1124264>
12. Park KH, Park KH, Kwon OY, Kang Y. A validity study of the Korean version of the interprofessional attitudes scale. *Korean Medical Education Review*. 2020;22(2):122-130. <https://doi.org/10.17496/kmer.2020.22.2.122>
13. Barnlund DC. A transactional model of communication. In: *Communication theory*. New York: Routledge; 2017. p. 47-57.
14. Kwon OY, Park KH, Park KH, Kang Y. Validity of the self-efficacy for interprofessional experimental learning scale in Korea. *Korean Medical Education Review*. 2019;21(3):155-161. <https://doi.org/10.17496/kmer.2019.21.3.155>
15. Comrey AL, Lee HB. *A first course in factor analysis*. New York: Psychology Press; 2013.
16. DeVellis RF, Thorpe CT. *Scale development: theory and applications*. Thousand Oaks: Sage Publications; 2021.
17. World Health Organization. WHODAS 2.0 translation guidelines [Internet]. Geneva (CH): World Health Organization; 2010 [cited 2024 Dec 9]. Available from: <https://terrance.who.int/mediacentre/data/WHODAS/Guidelines/WHODAS%202.0%20Translation%20guidelines.pdf>
18. Cai L, Choi K, Hansen M, Harrell L. Item response theory. *Annual Review of Statistics and Its Application*. 2016;3(1): 297-321. <https://doi.org/10.1146/annurev-statistics-041715-033702>
19. Chung H. The rasch model: an alternative method for analyzing ordinal data. *Journal of Coach Development*. 2005;7(3): 133-141.
20. Lee DY, Yang HJ, Yang DS, Choi JH, Park BS, Park JY. Rasch analysis of the clinimetric properties of the Korean dizziness handicap inventory in patients with parkinson disease. *Research in Vestibular Science*. 2018;17(4):152-159. <https://doi.org/10.21790/rvs.2018.17.4.152>
21. Linacre JM. Optimizing rating scale category effectiveness. *Journal of Applied Measurement*. 2002;3(1):85-106.
22. Asiedu K. Rasch analysis of the standard patient evaluation of eye dryness questionnaire. *Eye & Contact Lens: Science and Clinical Practice*. 2017;43(6):394-398. <https://doi.org/10.1097/ICL.0000000000000288>
23. Linacre JM. Rasch model estimation: further topics. *Journal of Applied Measurement*. 2004;5(1):95-110.
24. Seong TJ. *Understanding and application of item response theory*. Paju: Kyoyookbook; 2001.
25. Kim SH, Park JH. Development and validation of a tool for evaluating core competencies in nursing cancer patients on chemotherapy. *Journal of Korean Academy of Nursing*. 2012; 42(5):632-643. <https://doi.org/10.4040/jkan.2012.42.5.632>
26. Agbo AA. Cronbach's  $\alpha$ : review of limitations and associated recommendations. *Journal of Psychology in Africa*. 2010;20(2): 233-239. <https://doi.org/10.1080/14330237.2010.10820371>
27. Keiser MM, Turkelson C, Smith LM, Yorke AM. Using inter-professional simulation with telehealth to enhance teamwork and communication in home care. *Home Healthcare Now*. 2022;40(3):139-145. <https://doi.org/10.1097/NHH.0000000000001061>
28. Shim K, Shin H. The reliability and validity of the lasater clinical judgement rubric in Korean nursing students. *Child Health Nursing Research*. 2015;21(2):160-167. <https://doi.org/10.4094/chnr.2015.21.2.160>
29. Gwet KL. Computing inter-rater reliability and its variance in the presence of high agreement. *British Journal of Mathematical and Statistical Psychology*. 2008;61(1):29-48. <https://doi.org/10.1348/000711006X126600>
30. Terwee CB, Mokkink LB, Knol DL, Ostelo RW, Bouter LM, de Vet HC. Rating the methodological quality in systematic reviews of studies on measurement properties: a scoring system for the COSMIN checklist. *Quality of Life Research*. 2012;21: 651-657. <https://doi.org/10.1007/s11136-011-9960-1>
31. World Health Organization. *Patient safety curriculum guide: multi-professional edition*. Geneva: World Health Organization; 2011.
32. Salik I, Ashurst JV. *Closed loop communication training in medical simulation* [Internet]. Treasure Island (FL): StatPearls Publishing; 2023 [cited 2024 Dec 9]. Available from: <https://pubmed.ncbi.nlm.nih.gov/31751089/>
33. Paranduk R, Karisi Y. The Effectiveness of non-verbal communication in teaching and learning English: a systematic review. *Journal of English Education, Literature and Culture*. 2020;8 (2):140-154. <https://doi.org/10.53682/eclue.v8i2.1990>
34. Jun HS, Ju HJ. The effect of term based learning on communication ability, problem solving ability and self-directed learning in nursing science education. *Journal of Digital Convergence*. 2017;15(10):269-279. <https://doi.org/10.14400/JDC.2017.15.10.269>
35. Jang Y, Kim M, Lee J. IRTree model: an alternative approach for self-reported ordinal data analysis. *Journal of Educational Evaluation for Health Professions*. 2019;32(2):303-323. <https://doi.org/10.31158/JEEV.2019.32.2.303>