Validating the Korean shorter Diagnostic Thinking Inventory in medical education: a pilot study

Jihyun Si

Department of Medical Education, Dong-A University College of Medicine, Busan, Korea

Purpose: Developing clinical reasoning across the medical curriculum requires valid, reliable, and feasible assessment tools. However, few validated tools are available for the convenient and efficient quantification of clinical reasoning. Thus, this study aimed to create a shorter version of the Diagnostic Thinking Inventory (DTI) and validate it in the Korean medical education context (DTI-SK).

Methods: The DTI-SK was constructed using content validity and a translation and back-translation process. It comprises two subcategories and 14 items. Its validity and reliability were explored using exploratory and confirmatory factor analyses, mean comparisons of four medical student groups (med 1 to med 4), and internal consistency using Cronbach’s α. Two hundred medical students were invited to participate through email, and the survey was administered for 2 weeks.

Results: Data from 136 students were analyzed. Exploratory factor analysis revealed two factors with eigenvalues greater than 1.0 and they together explained 54.65% of the variance. Confirmatory factor analysis demonstrated that the model had acceptable level of fit and convergent validity. Discriminant validity was confirmed using heterotrait-monotrait criterion. Group comparisons demonstrated that the med 4 students showed significantly higher scores than the med 1 and 2 students. The inventory exhibited strong internal consistency for all items (Cronbach’s α=0.906).

Conclusion: The findings indicated that the DTI-SK is a reliable and valid tool for measuring medical students’ clinical reasoning in the context of Korean medical education.

Key Words: Diagnostic Thinking Inventory, Clinical reasoning assessment, Validation, Medical education

Introduction

Clinical reasoning refers to the cognitive processes in which clinicians collect and interpret data to diagnose and treat patients [1–3]. Clinical reasoning critically impacts clinicians’ performance and, ultimately, the quality of patient care [4,5]. Thus, acquiring clinical reasoning skills is an essential goal at every stage of undergraduate medical education [5,6]. The development of clinical reasoning across the medical curriculum requires valid, reliable, and feasible assessment for clinical reasoning [3], but despite its importance, the assessment of clinical reasoning abilities remains challenging [5–7].

A robust evaluation tool is essential for medical educators to develop and evaluate effective teaching interventions for clinical reasoning development. Different assessment methods, such as scenario-based multiple-choice questions, extended matching questions, concept maps, essays, key feature examinations, and script-
concordance tests, have been used to assess clinical reasoning in classroom settings [3]. However, these assessment methods can raise validity and reliability issues [3]. Furthermore, they require significant effort and time to develop and score. The relative lack of discussion of teaching methods for developing clinical reasoning may also be attributed to the limitations of valid assessment methods that can examine their effectiveness conveniently and efficiently [8].

Bordage et al. [1] developed the Diagnostic Thinking Inventory (DTI) to quantify the clinical reasoning abilities of medical students and physicians based on script theory. Script theory proposes that the combination of specific medical knowledge and how it is stored in memory is a prime determinant of clinical reasoning [9]. Clinicians gather, analyze, and evaluate medical information based on the knowledge base or illness scripts to arrive at a diagnosis [9]. The experts with highly developed illness scripts can conduct an inquiry into the information required for diagnosis, while novices with poorly developed illness scripts struggle to know what questions to ask, remember what they have been told, use the data meaningfully, gather further specific data, or even begin to start analyzing the data [10]. Based on this concept, Bordage et al. [1] developed the DTI, which has two subcategories: flexibility in thinking (FT) and evidence for structure in memory (SM). Bordage et al. [1] suggest that FT involves using multiple approaches to explore diagnostic possibilities based on key patient interview features or general inquiries when forceful features do not arise yet. In addition, SM refers to the availability and accessibility of organized knowledge stored in memory during clinical reasoning [11]. Bordage et al. [1] validated the DTI by demonstrating significant differences in the means scores between the first and third-year medical students and clinicians and a good level of internal consistency.

DTI is rooted in script theory; however, owing to its context-neutral design, this tool can assess clinical reasoning abilities in various clinical contexts. DTI has proven to be a valid and reliable tool in studies with medical students, medical residents, dental students, and physicians, as well as other health professionals who need clinical reasoning abilities to provide quality care, including optometrists, physiotherapists, athletic training, and nursing [7,8,11-19]. For example, Beullens et al. [12] explored the clinical reasoning of final-year medical students after problem-solving clinical seminars for 2 months. They measured the students’ clinical reasoning skills through extended matching questions and pre- and post-DTI. They reported that the students’ DTI scores significantly improved after the seminar, and the test’s scores (extended match questions) were significantly correlated with the DTI scores. In addition, the study of Rowat and Suneja [19] showed that after implementing a longitudinal clinical reasoning development curriculum, comparing the cohort with and without, the DTI of the second-year medical students (the cohort with) significantly improved. Kumar et al. [17] also reported that expert clinicians recognized by their peers showed higher mean scores than peer clinicians when their clinical reasoning skills were measured using the DTI in internal medicine.

Although DTI is a reliable tool for determining the level of FT and SM that experts and medical students use in clinical reasoning processes [1,7,12], it consists of 41 overlapping items that can be inconvenient to use. Edgar et al. [7] recently adapted DTI for use in optometric education contexts through context-based changes. They created a shorter version of the DTI in optometric education contexts (DTI-OS) by reducing items from the original DTI. The DTI-OS consists of 12 items, with six items for each subcategory. The DTI-OS in their study demonstrated a high level of internal reliability, with a
Cronbach's \( \alpha \) of 0.92, and its validity was established with significantly higher scores by qualified optometrists compared to second-year optometry students. Edgar et al. [7] also measured the clinical reasoning abilities of second-year optometry students using the original DTI. Despite the different contexts, its mean value (155.6) of second-year optometry students in their study was comparable to the mean value (158.3) of third-year medical students in Bordage et al. [1].

If it were possible to measure clinical reasoning abilities with a minimum number of DTI items, it could significantly reduce the time and effort required for assessment, enhance the tool's utility, and provide substantial assistance in clinical reasoning education. In addition, DTI has been used and validated in various contexts. However, validation studies related to DTI have not been conducted in the Korean medical education context.

Thus, the purpose of this pilot study was to create a shorter version of the DTI in medical education contexts based on the DTI-OS and validate it in the Korean medical context (DTI-SK). With few validated tools available to quantify clinical reasoning, the DTI-SK could be a convenient and valuable tool supporting development of clinical reasoning in the Korean medical education context.

**Methods**

1. **Participants and procedure**

Two hundred medical students from a private medical school in Korea were invited to participate via email (58 in the first year, 44 in the second year, 44 in the third year, and 54 in the fourth year). The medical school where this study was conducted has a curriculum that explicitly focuses on training clinical reasoning skills such as problem-based learning (PBL) (from the fourth quarter of the first-year program to the second quarter of the second-year program), clinical reasoning methods (during the third and fourth quarters of the second-year program), clinical clerkship and CPX (clinical performance examination) programs. When the survey was conducted at the beginning of the third quarter, first-year students did not take any explicitly structured clinical reasoning development programs, and second-year students took PBL from the fourth quarter of their first year to the second quarter of their second year. The survey included 44 items, including sex, age, and year, and was administered via email for 2 weeks. This study protocol was approved by the Institutional Review Board of Dong-A University (approval no., 2-1040709-AB-N-01-202109-DR-070-06). Informed consent was obtained from all participants before they participated in the survey.

2. **Content validity**

Content validity was explored by two professors of critical care medicine and allergy and clinical immunology, who have been conducting a Clinical Reasoning Method course for the past 3 years at a medical school, and two professors of emergency medicine and medical education, with research interests in clinical reasoning development. They were provided with the 12 items of DTI-OS and the original DTI through email and requested to review and comment on whether they thought that it measures clinical reasoning, whether it covers concepts related to clinical reasoning in medical education contexts, and to identify any items from the original DTI version that must be included in the DTI-SK. The four experts agreed that the inventory measures clinical reasoning and covers its concepts of clinical reasoning. However, one professor proposed including two items, item S10 (forceful...
feature, SM) and item F28 (comparing and contrasting diagnosis, FT), from the original DTI. Both were considered essential components of clinical reasoning; therefore, the DTI–SK was developed by incorporating these two items. Each subcategory consisted of seven items, resulting in 14 items.

3. Tools

The original DTI contains 41 items (FT, 21 items and SM, 20 items). The DTI–SK includes 14 items (seven in each subcategory), and each item consists of a stem followed by two statements that represent the opposite ends of a continuum on a 6-point scale, just like the DTI. The items alternate between positive statements on the right and left sides to avoid complacency. With appropriate correction of the reverse-scored items, a score was added, and a higher score represented more advanced clinical reasoning.

4. Translation

After receiving approval for DTI usage from the publisher (John Wiley and Sons), the procedures of translation and back-translation were conducted, following the guidelines for cross-cultural adaptation of self-reported measures [20,21]. To compare with the original DTI scores, a translation process was conducted for the entire DTI, including the DTI–SK. First, one bilingual individual, who had studied in an English-speaking country for more than 10 years and was fluent in English and Korean, translated the items into Korean. Second, one emergency medicine doctor proficient in Korean and English reviewed the items. He made some semantic changes to ensure semantic and content equivalence. For example, in the item 15, “I am quite happy to dismiss some information as irreverent” was translated as “I simply ignore some information as irrelevant.” Third, one individual, blinded to the original English version, had studied in an English-speaking country for more than 5 years and who was proficient in English and Korean, back-translated the items from Korean to English. Fourth, the back-translated items were reviewed by the researcher and translator together and compared with the original tool. Minor inconsistencies were found and subsequently corrected. Finally, professors of critical care and emergency medicine reviewed the revised Korean version. One item was refined based on their feedback. Then, the tool was subjected to a preliminary investigation involving three first-year medical students; after incorporating their feedback, final adjustments were made to clarify unclear items, and the tool was completed. The Korean version of a shorter DTI (DTI–SK) was presented in Supplement 1.

5. Statistical analysis

Statistical analyses were conducted using the IBM SPSS Statistics (IBM Corp., Armonk, USA) and AMOS ver. 27.0 (IBM Corp.). Participant characteristics were analyzed using frequency analysis. The construct validity of the DTI–SK was evaluated using individual item analysis, factor analysis, heterotrait–monotrait (HTMT) ratio of correlations, and mean comparisons of the four student groups (med 1, med 2, med 3, and med 4). For individual item analysis, descriptive statistics and corrected item-total correlations were calculated. Factor analysis was undertaken through two sequential steps involving exploratory and confirmatory factor analyses. Exploratory factor analysis involved extracting two factors using principal component analysis and varimax rotation, with an eigenvalue of 1.0. The specified factor model was subsequently validated through confirmatory factor analysis, which assessed model fit and convergent validity. Discriminant validity was evaluated using HTMT criterion proposed by Henseler et al. [22]. The HTMT value was calculated using the HTMT online calculator developed by Henseler et al. [22] (https://www.henseler.com/htmt.html).
For mean comparisons among student groups, an analysis of variance (ANOVA) and post-hoc analyses using Tukey’s tests were conducted. For comparison with DTI-SK, the original DTI scores were also analyzed among the groups using ANOVA and post-hoc analyses using Tukey’s tests. The reliability was verified using Cronbach’s α. A p-value less than 0.05 was considered statistically significant.

Results

1. Participants’ characteristics

Of the 200 students invited to participate in the survey, 151 students (78.5%) answered the survey. Among the 151 students, five students did not complete the survey, and eight students consistently chose either 1 or 6 for all questions or alternated between 1 and 6. Excluding these 15 students, which includes two outliers from the fourth-year student group, data of 136 participants were analyzed. Sample sizes for factor analysis are considered adequate when there are roughly 200 cases or a ratio of 5:1 between cases and measurement variables [23]. In this study, the ratio of cases to measurement variables is 136:14, which exceeds the 5:1 threshold. The final analysis included 42 med 1, 38 med 2, 15 med 3, and 41 med 4 students. Among the participating students, there were 52 female students (38.2%), and 84 male students (61.8%). Their average age was 24.41 years (standard deviation= 2.10), with an age range of 20-35 years.

2. Individual item analysis

The mean, standard deviation, corrected item-total correlation, skewness, and kurtosis of each DTI-SK item were analyzed. The results are shown in Table 1. The normality criteria for skewness and kurtosis for each item fell below the absolute values of 3 and 8, respectively, indicating that all items met the conditions for a normal distribution [24]. In addition, the corrected item-total correlations ranged from 0.50-0.70, with all items meeting the recommended correlation threshold of 0.4 or higher for factor analysis [25].

Table 1. Results of Individual Item Analysis, Exploratory Factor Analysis, and Internal Consistency Analysis

<table>
<thead>
<tr>
<th>Item</th>
<th>Factor 1</th>
<th>Factor 2</th>
<th>Mean ± SD</th>
<th>ITC</th>
<th>Skewness</th>
<th>Kurtosis</th>
<th>α if removed</th>
<th>α</th>
</tr>
</thead>
<tbody>
<tr>
<td>S13</td>
<td>0.784</td>
<td></td>
<td>4.01±1.04</td>
<td>0.61</td>
<td>-0.83</td>
<td>0.17</td>
<td>0.910</td>
<td>0.870</td>
</tr>
<tr>
<td>S10</td>
<td>0.678</td>
<td></td>
<td>4.23±1.06</td>
<td>0.50</td>
<td>-0.55</td>
<td>0.002</td>
<td>0.904</td>
<td></td>
</tr>
<tr>
<td>S33</td>
<td>0.677</td>
<td></td>
<td>4.11±1.05</td>
<td>0.62</td>
<td>-0.53</td>
<td>-0.13</td>
<td>0.899</td>
<td></td>
</tr>
<tr>
<td>S9</td>
<td>0.676</td>
<td></td>
<td>4.22±1.13</td>
<td>0.70</td>
<td>-0.45</td>
<td>-0.25</td>
<td>0.896</td>
<td></td>
</tr>
<tr>
<td>S19</td>
<td>0.675</td>
<td></td>
<td>3.59±1.05</td>
<td>0.64</td>
<td>-0.16</td>
<td>-0.40</td>
<td>0.898</td>
<td></td>
</tr>
<tr>
<td>S8</td>
<td>0.646</td>
<td></td>
<td>4.01±1.17</td>
<td>0.66</td>
<td>-0.34</td>
<td>-0.58</td>
<td>0.897</td>
<td></td>
</tr>
<tr>
<td>S31</td>
<td>0.561</td>
<td></td>
<td>3.74±0.97</td>
<td>0.66</td>
<td>-0.16</td>
<td>-0.36</td>
<td>0.898</td>
<td></td>
</tr>
<tr>
<td>F38</td>
<td>0.742</td>
<td></td>
<td>3.62±0.97</td>
<td>0.55</td>
<td>-0.11</td>
<td>0.34</td>
<td>0.902</td>
<td>0.844</td>
</tr>
<tr>
<td>F15</td>
<td>0.710</td>
<td></td>
<td>3.57±1.15</td>
<td>0.56</td>
<td>-0.02</td>
<td>-0.79</td>
<td>0.902</td>
<td></td>
</tr>
<tr>
<td>F4</td>
<td>0.703</td>
<td></td>
<td>3.71±0.99</td>
<td>0.66</td>
<td>-0.24</td>
<td>0.004</td>
<td>0.890</td>
<td></td>
</tr>
<tr>
<td>F27</td>
<td>0.703</td>
<td></td>
<td>3.43±1.02</td>
<td>0.57</td>
<td>0.08</td>
<td>-0.44</td>
<td>0.901</td>
<td></td>
</tr>
<tr>
<td>F41</td>
<td>0.692</td>
<td></td>
<td>4.14±1.05</td>
<td>0.66</td>
<td>-0.75</td>
<td>0.82</td>
<td>0.898</td>
<td></td>
</tr>
<tr>
<td>F28</td>
<td>0.652</td>
<td></td>
<td>3.72±1.07</td>
<td>0.56</td>
<td>0.05</td>
<td>-0.35</td>
<td>0.901</td>
<td></td>
</tr>
<tr>
<td>F5</td>
<td>0.647</td>
<td></td>
<td>3.82±1.11</td>
<td>0.56</td>
<td>-0.35</td>
<td>-0.73</td>
<td>0.902</td>
<td></td>
</tr>
<tr>
<td>EV</td>
<td>3.878</td>
<td></td>
<td>3.773</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EXV</td>
<td>27.70%</td>
<td></td>
<td>26.95%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CV</td>
<td>54.65%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>


*The item numbers match the item numbers in the original Diagnostic Thinking Inventory.
3. Exploratory factor analyses

The Kaiser–Meyer–Olkin (KMO) and Bartlett’s sphericity tests were conducted to assess the adequacy of the factor analysis. The KMO index yielded a value of 0.917, indicating that the sampling adequacy was excellent [26]. Additionally, Bartlett’s sphericity test resulted in a chi-square value of 803.050 (degrees of freedom [df]=91, p<0.000), indicating that the data demonstrated the value of conducting factor analysis. Subsequently, factor analysis was conducted and yielded two factors with eigenvalues greater than 1.0. The first factor explained 27.70% of the variance and the second explained 26.95%. These two factors together accounted for 54.65% of the variance, meeting the recommended threshold for explained variance of 0.50%–0.60% [26]. The factor loadings for all the items were above 0.50, ranging from 0.561 to 0.764 (Table 1).

4. Confirmatory factor analyses

To determine how well the specified factor model represented the data, various goodness-of-fit indices were examined. The model yielded an acceptable level of fit [27]: chi-square divided by degree of freedom=1.096 (df=76, p=0.265), standardized root mean squared residual =0.043, root mean square error of approximation=0.027, comparative fit index=0.990, and normed fit index=0.901 (Table 2). Convergent validity was verified with standardized factor loadings, their significance level, and composite reliability (CR). All variables exhibited standardized factor loadings surpassing 0.50, the recommended threshold [25,28] with all t-values being significant (p<0.001). In addition, CR exceeded the 0.70 threshold as seen Table 3, indicating a high level of reliability. Also, for convergent validity, the criterion by Fornell and Larcker [29] suggested that average amount of variance (AVE) should be equal or greater than 0.50. However, Malhotra and Dash [30] argue that AVE is often too strict, and reliability can be established through CR alone. In this study, CR exceeded 0.8 and surpassed AVE, so convergent validity of the construct is thought to be adequate. Discriminant validity was evaluated using HTMT criterion proposed by Henseler et al. [22]. The HTMT value for each pair of constructs, computed on the basis of the item correlations, exceeds 0.85, indicating discriminant validity problems. As seen in Table 3, the HTMT value in this study was 0.784, confirming discriminant validity.

5. Group comparisons

ANOVA was conducted to compare the four groups, and the results are presented in Table 4. Significant differences were observed in the overall DTI-SK scores and the two subcategories. Post-hoc analyses using Tukey’s tests demonstrated significant differences between the med1

---

Table 2. Model Fit Summary

<table>
<thead>
<tr>
<th>Indicator (acceptable level)</th>
<th>CMIN/DF (&lt;3.0)</th>
<th>p-value</th>
<th>SRMR (&lt;0.08)</th>
<th>RMSEA (&lt;0.08)</th>
<th>CFI (&gt;0.90)</th>
<th>NFI (&gt;0.90)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Value</td>
<td>1.906 (df=76)</td>
<td>0.265</td>
<td>0.043</td>
<td>0.027</td>
<td>0.990</td>
<td>0.901</td>
</tr>
</tbody>
</table>

CMIN/DF: Chi-square divided by degree of freedom, SRMR: Standardized root mean squared residual, RMSEA: Root mean square error of approximation, CFI: Comparative fit index, NFI: Normed fit index.

Table 3. Model Validity Analysis

<table>
<thead>
<tr>
<th></th>
<th>CR</th>
<th>AVE</th>
<th>Correlation</th>
<th>HTMT</th>
</tr>
</thead>
<tbody>
<tr>
<td>FT</td>
<td>0.846</td>
<td>0.440</td>
<td>0.791***</td>
<td>0.784</td>
</tr>
<tr>
<td>SM</td>
<td>0.871</td>
<td>0.492</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

and med 4 groups in both overall DTI-SK, FT, and SM scores and between the med 2 and med 4 groups in both overall DTI-SK and SM scores. In addition, the original version of the DTI scores was analyzed using ANOVA. Significant differences were found in both overall DTI and the two subcategories. Post-hoc analyses using Tukey’s tests demonstrated significant differences between the med 1 and med 4 groups in both the overall DTI, FT, and SM scores and between the med 2 and med 4 groups in both overall DTI and SM scores.

6. Internal consistency

The internal consistency of the DTI-SK was explored using Cronbach’s α; it was 0.906 for all items, 0.844 for the FT, and 0.870 for the SM. These values indicated a good level of internal consistency. No items were found to increase the reliability coefficient when removed from the analysis (Table 1).

Discussion

This study created and validated DTI-SK in the context of Korean medical education. Its content validity was established by a group of experts, and it exhibited excellent internal consistency with a Cronbach’s α of 0.906. The FT and SM subcategories also demonstrated good internal consistency, with values of 0.870 and 0.844, respectively. These values are, in general, slightly lower than those reported by Edgar et al. [7] (0.92 for the overall score, 0.90 for FT, and 0.83 for SM), but higher than those reported by Bordage et al. [1] (0.83 for the overall score, 0.72 for FT, and 0.74 for SM). The inventory as a whole gave an alpha value above 0.9, implying that the items measured the same thing.

Construct validity was evaluated using exploratory and confirmatory factor analysis and mean comparisons among student groups. Exploratory factor analysis extracted two factors, FT and SM, which are consistent with previous
research findings. All 14 items demonstrated factor loadings of 0.05 or higher, and their corrected item-total correlation coefficients were all above 0.40. In addition, the specified factor model fitted very well with the collected data and the relationships between the observed variables and latent variables were significant. Although the AVE values for both FT and SM did not reach 0.05, the CR exceeded 0.80 and the HTMT value between two latent constructs was below the stringent threshold of 0.85. Consequently, these results are considered to provided additional supports for the validity and reliability of the constructs underlying the developed tool.

It was hypothesized that a valid tool would yield higher scores in the med 4 student group who had almost completed their preparation for the national practical examination for medical practitioners and had begun taking the practical exam, compared to the med 1 student group. As expected, in the comparison of groups, the med 4 student group scored significantly higher on both the overall DTI-SK and its subcategories when compared to the med 1 and med 2 student groups, which demonstrated construct validity. These results are similar to those reported by the study of Edgar et al. [7], demonstrating significant differences between qualified optometrists and second-year optometry student groups by measuring their clinical reasoning abilities using DTI-OS.

This pattern was also observed in the original DTI scores. The med 4 students scored higher than the med 1 students in both the overall DTI, FT, and SM, and they also scored higher than the med two students in overall DTI and SM. This result aligns with those of previous studies. In the studies by Bordage et al. [1] and Edgar et al. [7], the comparison between the student groups and the groups with qualification demonstrated significant differences in the overall DTI scores. In addition, in this study, the overall DTI score of the third-year medical students was 156.13, with FT and SM scores of 76.33 and 79.80, respectively. These scores are comparable to those in the study by Bordage et al. [1], in which the total DTI score of third-year medical students was 158.3, with FT and SM scores of 81.6 and 76.7, respectively, and those in the study by Edgar et al. [7], which the total DTI score of second-year optometry students was 155.62, with FT and SM scores of 79.13 and 76.45, respectively.

One critical educational goal of medical schools is to develop medical students’ clinical reasoning abilities; however, it is challenging to teach and even more challenging to assess. There is a growing need for tools with established reliability and validity to conveniently and efficiently evaluate clinical reasoning abilities. The results of this study indicate that the DTI-SK, comprising two subcategories and 14 items, can be used as a valid and reliable tool to assess medical students’ clinical reasoning abilities in the context of Korean medical education. Considering that multiple measurements are often conducted to assess clinical reasoning abilities after applying interventions to improve them, the DTI-SK can be a convenient and efficient tool for measuring clinical reasoning abilities with a smaller number of items.

The DTI-SK demonstrated a clear distinction in clinical reasoning abilities. Thus, learners can use this tool to self-diagnose their strengths and weaknesses regarding FT and SM and gain insights into areas where further learning is required [1,7,15]. It is interesting to note that the SM scores were higher than the FT scores in this study. This is understandable, as the students did not have much clinical experience, and the FT skills may come through experience with managing patient cases [7,13,14]. This result indicates that more time is needed to train medical students for flexibility in the thinking for patient interviews.

In conclusion, the findings indicate that the DTI-SK is a valid and reliable inventory for measuring medical students’ clinical reasoning in Korean medical education.
In addition, it can be applied to various areas of clinical practice owing to its context-neutral design. With the enhancement of the ease and quality of clinical reasoning assessment, it can facilitate and enhance the teaching and assessment of clinical reasoning in medical students.

This study had several limitations. This pilot study was limited to a specific geographical area and university and may not fully represent the entire population. This study did not include data from qualified clinicians. Future studies should compare medical students and qualified experts to further confirm the validity of DTI-SK. Additionally, this study did not conduct criterion-related validity checks due to the lack of widely accepted valid tools for assessing clinical reasoning. Instead, this study presents the original DTI scores for comparisons. Furthermore, the AVE values in this study did not reach 0.05, which could be a limitation in terms of discriminant validity. Although the HTMT value indicated acceptable discriminant validity, further study is needed to examine this issue. Finally, DTI-SK is a self-reported survey; therefore, it is subject to social desirability bias.

Supplementary materials
Supplementary files are available from https://doi.org/10.3946/kjme.2024.281

Supplement 1. DTI-SK: The Korean Shorter Diagnostic Thinking Inventory.

Author contributions: Jihyun Si designed the study, collected and analyzed data, wrote and revised the manuscript.

References

10. Schmidt HG, Mamede S. How to improve the teaching of clinical reasoning: a narrative review and a proposal.


