EUROPEAN REHABILITATION JOURNAL | ORIGINAL

Development and validation of a script concordance test to assess clinical reasoning in physiotherapy students focusing on lower limb tendinopathies

Jérémy Vanderweyen, PT, [1,2], **Gérard Ndacayisaba, PT,** [1,2] **and Joachim Van Cant, PT, PhD,** [1,2]
[1]Unité de Recherche en Sciences de la Réadaptation, Bruxelles,Belgium, [2]Faculté des Sciences de la Motricité Humaine, Université Libre de Bruxelles, Belgium

received : 09 January 2025
accepted: 23 January 2026

ISSN: 2823-989X
DOI: 10.52057/erj.v6i1.72

ABSTRACT

Background: Tendinopathies are prevalent, complex musculoskeletal conditions that challenge clinical decision-making. While various tools have been developed to assess clinical reasoning across health professions, the Script Concordance Test (SCT) has shown promise as a context-sensitive method, though its application in physiotherapy particularly for tendinopathy has not been explored. Objective: This study aimed to develop and assess the reliability and validity of an SCT administered to final-year physiotherapy students and experts, focusing on the diagnosis and treatment of lower limb tendinopathies. **Method**: Three experts in musculoskeletal disorders developed the script concordance test (SCT) with a structured three-phase process. They created 16 clinical scenarios focused on lower limb tendinopathies, resulting in a preliminary version of the test composed of 48 items. These items were administered online to a panel of 15 expert physiotherapists and 50 final-year physiotherapy students from the Université Libre de Bruxelles. **Results**: After optimization, the SCT demonstrated an internal consistency, reflected by a Cronbach's alpha coefficient of 0.69. A statistically highly significant difference (P<0.001) was noted between the two groups. The average score for students was 61.3 (±8.4), whereas experts scored 74.6 (±6.4). **Conclusion**: The results suggest that our SCT sequence is valid and reliable when it comes to distinguishing clinical reasoning levels between students and experts from our sample on the management of lower limb tendinopathies. This study highlights the potential of a context-specific SCT sequence to assess clinical reasoning in physiotherapy students and experts managing lower limb tendinopathies.

KEYWORDS: Clinical reasoning, Lower limb tendinopathies, Physiotherapy, Script concordance test

Background

Tendinopathies, which are characterized by tendon degeneration and pain, frequently resulting from overuse or mechanical overload [1], represent a significant burden on both society and the healthcare system. Despite advancements in our understanding of these conditions, their management remains a considerable challenge for healthcare professionals. Tendinopathies and tendon tears account for approximately 30% of consultations related to musculoskeletal disorders [2]. Lower limb tendinopathies are particularly prevalent among both athletes and the general population and are frequently associated with pain and functional limitations [3, 4]. These conditions are often difficult to treat and may negatively impact sports participation, work capacity, and overall quality of life [5, 6]. Their complex clinical presentations, combined with the absence of standardized treatment protocols, further complicate management [7]. Physiotherapy plays a central role in the treatment of tendinopathies, with well-established interventions aiming to alleviate pain, restore function, and improve quality of life. However, diagnosis and management remain challenging due to the variability of clinical presentations and ongoing debate regarding optimal treatment strategies [8].

Clinical reasoning is a fundamental component of clinical competence and is essential for delivering effective and efficient care [9]. It encompasses the cognitive processes and decision-making skills required in clinical contexts [10]. Numerous tools exist for assessing clinical reasoning, al-

Corresponding author:
Jérémy Vanderweyen, PT, Route de Lennik 808. B-1070 Bruxelles, Belgique. e-mail: jeremy.vanderweyen@ulb.be

though they vary in design, scope, and applicability across disciplines and specialties [11]. These assessments may be employed in both workplace-based and non-workplace-based settings [11]. Among the latter, the Script Concordance Test (SCT) has emerged as a valuable tool. SCTs consist of brief clinical vignettes describing scenarios marked by uncertainty, which learners may realistically encounter. Each vignette presents a clinical hypothesis, followed by the introduction of new information. This design encourages students to process and integrate new information within the context of existing knowledge, simulating real-life clinical decision-making where ambiguity is common, and information is often revealed incrementally. The SCT format assesses learners' ability to mobilize and apply organized knowledge networks to resolve complex clinical problems. These scenarios are intentionally complex even for experts and typically present multiple plausible diagnostic or therapeutic options. The tool is grounded in script theory, which conceptualizes clinical knowledge as structured networks developed through theoretical learning and clinical experience. SCTs measure the degree of concordance between a learner's reasoning and that of a panel of experts, thereby reflecting the organization and application of their clinical knowledge [12].

Although SCTs have been developed in several health professions including medicine, nursing, optometry, and dentistry their use in physiotherapy education remains limited [13]. In physiotherapy, SCTs have mostly been explored in domains such as stroke rehabilitation and assistive device prescription [14]. These tools are typically developed in collaboration with experienced clinicians to ensure content validity and clinical relevance. However, no SCT has yet been developed to assess clinical reasoning in the management of lower limb tendinopathies. To address this gap, the present study aims to develop and explore the validity of a SCT designed to assess clinical reasoning in the context of lower limb tendinopathy management in physiotherapy. This tool may help identify reasoning patterns among students and contribute to the improvement of educational strategies in this area of clinical training.

Material and methods

Population

We used convenience sampling for this study. Recruitment of subjects was achieved through social networks. Our goal was to form two groups of participants: a group of experts and a group of final-year physiotherapy students from the Université Libre de Bruxelles. For the expert group, the criteria included french-speaking expert physiotherapists from France and Belgium who had completed continuing education in manual therapy accredited by the International Federation of Orthopaedic Manipulative Physical Therapists (IFOMPT). This accreditation ensures high standards of training and practice in the field of musculoskeletal disorders [15]. Additionally, these participants were required to have at least three years of field experience. For the student group, the criteria included french-speaking students from France and Belgium enrolled final-year of physiotherapy at the Université Libre de Bruxelles. The enrolled students were native French speakers, which ensured an adequate understanding of the test written in French. Fifteen experts were recruited to form the expert panel (Group 1), and all completed the test. Responses were also obtained from 50 participants in the student group (Group 2). For SCTs, a panel of 10–20 members is recommended to achieve adequate precision in terms of estimated reliability [16]. Other studies recommended 15 experts, the number we aimed for [17, 18].

Development of the script concordance test

The development of the test was carried out in three steps, as illustrated in Figure 1. To guide the development of the test, a specification table was first established to define its overall structure and ensure balanced item distribution across two domains: diagnosis and treatment. This framework was used to guarantee content validity and to align the test content with clinical situations commonly encountered by physiotherapy students during training and in professional practice. The test focused on four types of lower limb tendinopathies: Achilles, patellar, hamstring, and gluteus medius. For each condition, four distinct clinical scenarios were developed, resulting in a total of 16 clinical cases. Each case included three items, leading to a final test composed of 48 items. Each scenario presented a diagnostic or treatment hypothesis, followed by the introduction of new information. Respondents then use a Likert scale (−2 to +2) to indicate the impact of this information on the initial hypothesis [19]. Patient profiles varied in age and gender to reflect the diversity of real-world clinical cases. Clinical cases were written by a physiotherapist and investigator, (JV) specialized in musculoskeletal pathology management, manual therapist and who is familiar with clinical reasoning pedagogical approach. Secondly, other specialists (JVC, GN) in the management of musculoskeletal disorders reviewed the vignettes, leading to the initial modification of the test. This review involved reformulating certain clinical situations that were deemed inadequate. The investigators ensured that the test content aligned with the master's program in physiotherapy at the Université Libre de Bruxelles. Thirdly, the first version of the test was set up on the "Monkey Survey ©" platform. Students and experts enrolled in the study were provided with a link to access the online questionnaire. All subjects responded to the initial 48 items. The test was then optimized by analyzing the responses from the 15 experts, identifying problematic items. These items were identified using a Canadian SCT score calculator from the Université de Montréal. The final version of the test retained 42 items. Its process allowed us to improve validity and reliability of our TCS.

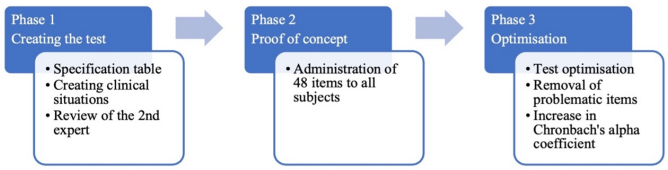

Figure 1 Steps in the construction of the SCT

Scoring the test

SCT questions were designed to avoid having single 'correct' or 'consensus' answers. Instead, scoring of the SCT was based on an aggregate method that considered the observed variability of responses of experts to clinical situations. The scoring system of the SCT was designed to compare a student's clinical reasoning with that of a group of experts. The key idea behind the scoring method is that there is not one single correct answer. Instead, the answer considered "best" is the one chosen by most of the expert panel, also known as the modal response (Table 1). This unanimous or majority consensus represents an expert's way of reasoning under uncertainty. For each question, if a student selected the modal response, they received a full credit of 1 point for that item. In other words, the student's reasoning closely matched the most common expert opinion. If the student chose a response that fewer experts supported, then they were awarded partial credit. Partial credit was calculated by taking the number of experts who selected that particular response and dividing it by the number of experts who chose the modal response. For example, in a scenario where 8 out of 15 experts selected a particular option, a less common response might be credited as, say, 5 divided by 8 (if 5 experts had given that answer), which would result in a score of 0.625 for that item. The scoring process for the experts involves analyzing each answer chosen by an expert for each item. This process involved subtracting 1 from the total number of experts who chose the same option and then dividing this number by the total number of experts who opted for the

modal response minus one. For example, consider an item answered by 15 experts. If the modal response (i.e., the most frequently selected option) was chosen by 8 experts, the maximum possible credit for that item was assigned to any expert who selected that same response. The score was calculated as (8−1) ÷ (8−1) = 1.00. An expert selecting a less common response for instance, one chosen by 5 experts would receive a partial credit of (5−1) ÷ (8−1) = 0.571. Similarly, a response selected by only 2 experts yielded a score of (2−1) ÷ (8−1) = 0.143. Responses not selected by any expert received a score of zero. This method ensures that each expert's answer was weighted based on its level of agreement with the reference panel, maintaining consistency with the SCT scoring philosophy that values reasoning convergence over absolute correctness (17). Descriptive statistics were calculated to summarize the SCT scores for both the expert and student groups. To assess the assumption of normality, a Shapiro-Wilk test was applied to the score distributions of each group. As the data followed a normal distribution, an independent samples Student's t-test was conducted to compare the mean SCT scores between groups. Statistical significance was set at $p < 0.05$. All analyses were performed using Jamovi version 2.3 (The Jamovi Project, 2023).

Table 1 Illustrative scoring table for students and experts

Selected Response	Number of experts	Student scoring	Expert scoring
−2	0	0	0
−1	2	2/8 = 0,25	(2-1) / (8-1) = 0,143
0	5	5/8 = 0,625	(5−1) / (8−1) = 0,571
+1	**8 (modal)**	8/8 = 1	(8−1) / (8−1) = 1
+2	0	0	0

Results

Socio-demographic analysis

Our sample included both male and female participants. We found that one-third of the expert panel had between eleven and twenty years of experience. Over two-thirds of the expert panel were between the ages of 24 and 40, with most of them being in their thirties. Most students were between the ages of 21 and 30.

Item selection and statistical analysis

At the end of our preliminary analysis, 6 items were removed because their correlation coefficient was too low ($r < 0.05$). The items that were excluded from the statistical analysis were items 6, 5, 9, 11, 28, and 36. Removing these low-correlation items increased the Cronbach's alpha coefficient. Following this item reduction process, the Cronbach's alpha coefficient went from 0.63 to 0.69. Cronbach's alpha was used to assess the internal consistency of the test. Internal consistency reflects the extent to which the items of an instrument measure the same underlying construct, in this case clinical reasoning. In Script Concordance Tests (SCTs), multiple items explore different aspects of reasoning in uncertain clinical situations. Therefore, an acceptable level of internal consistency suggests that the items collectively contribute to the measurement of the same construct. This increase suggests that the test's reliability improved because of removing the six poorly correlated items. Ultimately, 42 items were retained for the statistical analysis: 22 items in the diagnostic category and 20 items in the intervention category. This refined set of items ensured a more consistent measure of the clinical reasoning abilities of final-year physiotherapy students.

SCT scores

The analysis of the scores for students and experts revealed notable differences. The average score for students was 61.26%, with a median of 61.5%, a standard deviation of 8.5%. In contrast, experts had a higher average score of 74.8%, a median of 77%, a lower standard deviation of 6.5%. These results indicate that experts not only perform better on average but also show more consistency in their scores compared to students (Figure 2). Figure 2 graphically illustrates the distribution of the SCT scores of the two groups after optimization. The box plot for the student group shows a centered distribution relative to the total number of students, suggesting a balanced average. We also observe that the interquartile range (IQR) of the SCT scores ranged from 56% (Q1) to 68% (Q3). Prior to comparing the scores between the two groups, the normality of the data distributions was assessed. A Shapiro-Wilk test was conducted to assess the normality of the data distribution. The results indicated a W statistic of 0.968 and a p-value of 0.657. Since the p-value was greater than the conventional threshold of 0.05, we considered the distribution as normal. Because of the normal distribution, we used a t-test that revealed a statistically highly significant difference between the scores of the student and expert groups ($p < 0.001$).

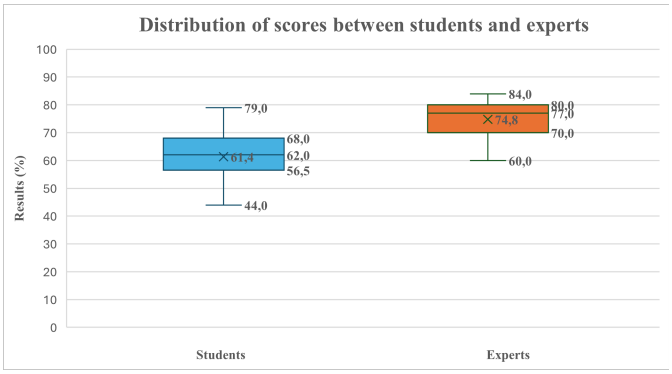

Figure 2 Boxplot representation of student and expert panel scores

Discussion

The objective of our study was to compare the clinical reasoning abilities of final-year physiotherapy students with a panel of experts. To achieve this objective, we developed a SCT specifically for the diagnosis and treatment of lower limb tendinopathy. Initially comprising 48 items, the test underwent optimization where 6 items were removed due to low correlation coefficients ($r < 0.05$). Ultimately, our test consisted of 42 items, with 20 in the intervention category and 22 in the diagnostic category.

The SCT demonstrated acceptable internal consistence, with a Cronbach's alpha coefficient of 0.69, which aligns with literature findings where coefficients range from 0.65 to 0.90 [20, 21, 22, 23, 24, 25]. As shown previously, our results revealed a statistically significant difference between the scores of students and experts ($p < 0.001$), demonstrating the test's ability to distinguish individuals with varying levels of expertise. Furthermore, the standard deviation within each group indicates that the clinical reasoning skills of students exhibited greater variability, while experts demonstrated a stronger consensus in their clinical reasoning.

The findings highlight the importance of ongoing training and professional experience in enhancing clinical reasoning capabilities when managing and diagnosing musculoskeletal disorders. The treatment perspectives and diagnosis of various tendinopathies are still under study today. Therefore, it is necessary to enhance clinical reasoning abilities, which are central to achieving effective and efficient clinical outcomes [9]. The

need for such tools is underscored by the limitations of traditional assessment methods, such as multiple-choice questionnaires and Objective Structured Clinical Examinations (OSCEs), which may not effectively evaluate clinical reasoning abilities in complex clinical contexts involving uncertainty [26]. Kojich and colleagues have pointed out the lack of correlation between SCT scores and other assessment methods, indicating that SCTs primarily assess the cognitive components of clinical reasoning, whereas OSCEs measure a broader range of skills, including cognitive, psychomotor, and affective domains [13].

In addition to their evaluative capabilities, SCTs can serve as valuable learning tools. This approach, referred to as concordance learning, leverages the complexity and uncertainty of real clinical situations and promotes pedagogical reflection [27, 28, 29]. Furthermore, using SCTs for educational purposes can help students pinpoint their weaknesses, fostering a desire for improvement [14].

Limitations of the Study

Despite the strengths of our study, some limitations should be acknowledged. One significant limitation is the relatively small size of our sample, which was affected by challenges in recruiting experts in the management of musculoskeletal disorders. Additionally, while the Cronbach's alpha coefficient for our test (0.69) aligns with existing literature, some studies suggest that a minimum of 0.70 is preferable for this coefficient to be deemed acceptable [16, 30]. A larger number of items could potentially increase this threshold. However, we aimed to avoid making the test excessively long, as this might discourage participants from completing it in its entirety.

Conclusion

Although the SCT is used in other health domains, few studies have employed it to assess clinical reasoning abilities in physiotherapy [14]. To our knowledge, our study is the first to compare the clinical reasoning of final-year physical therapy students with that of experts in the context of diagnosing and treating lower limb tendinopathies. Managing these pathologies can be complex and requires a multifactorial approach. The results indicated a statistically significant difference between the two groups, with the expert group achieving a higher average score. The SCT sequence we developed in this study demonstrated acceptable validity and reliability. In an era where physiotherapists are taking on increasing responsibilities, it is crucial to evaluate their clinical reasoning capacity, a cornerstone of effective patient care. Further studies are needed to develop and validate script concordance test sequences tailored to physiotherapy contexts. Such tools may be used both as formative instruments to support the development of clinical reasoning skills, and as formal assessment tools to certify competencies related to clinical reasoning within physiotherapy education programs.

Statement and declaration

Authors' contribution

JV, GN and JVC conducted data collection, conducted data analysis, interpreted findings. JV wrote the initial draft of the manuscript, which was then revised and approved by GN and JVC.

Data availability statement

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

Competing Interests

The authors declare they have no conflicting interests with the content of the article.

Ethics

The current study was accepted by ethical committee of Erasme hospital (ref: P2024/047).

Patient and public involvement

Patient and public involvement: Patients and/or the public were involved in the design, or conduct, or reporting, or dissemination plans of this research. Refer to the Methods section for further details.

Funding

The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

References

[1] Tomás F Fernández Jaén, Fernando Baró Pazos, Ascensión Fernández Jiménez, Marta Guillén Vicente, and Pedro Guillén García. Current concepts of the pathophysiology of tendinopathies. tissue engineering. *Apunts Sports Medicine*, 45(168):259–264, 2010.

[2] Richard R McCormack Jr, Robert D Inman, Athol Wells, Christopher Berntsen, and Harold R Imbus. Prevalence of tendinitis and related disorders of the upper extremity in a manufacturing workforce. *The Journal of Rheumatology*, 17(7):958–964, 1990.

[3] Suzan de Jonge, C Van Den Berg, Robert-Jan de Vos, HJL Van Der Heide, Adam Weir, JAN Verhaar, SMA Bierma-Zeinstra, and JL Tol. Incidence of midportion Achilles tendinopathy in the general population. *British Journal of Sports Medicine*, 45(13):1026–1028, 2011.

[4] Christer Rolf and Tomas Movin. Etiology, histopathology, and outcome of surgery in achillodynia. *Foot & Ankle International*, 18(9): 565–569, 1997.

[5] Angela Fearon, Jane E Dahlstrom, Jane Twin, Jill Cook, and Alex Scott. The Bonar score revisited: region of evaluation significantly influences the standardized assessment of tendon degeneration. *Journal of Science and Medicine in Sport*, 17(4):346–350, 2014.

[6] Henk Van der Worp, Mathijs Van Ark, Saskia Roerink, Gert-Jan Pepping, Inge Van Den Akker-Scheek, and Johannes Zwerver. Risk factors for patellar tendinopathy: a systematic review of the literature. *British Journal of Sports Medicine*, 45(5):446–452, 2011.

[7] Mahamed Ateef. Physiotherapy management in common tendon injuries: review of reviews. *International Journal of Physiotherapy*, 5 (2):69–74, 2018.

[8] Ashley Titan and Nelly Andarawis-Puri. Tendinopathy: investigating the intersection of clinical and animal research to identify progress and hurdles in the field. *JBJS Reviews*, 4(10):1–22, 2016.

[9] Nicole Christensen, Lisa Black, Jennifer Furze, Karen Huhn, Ann Vendrely, and Susan Wainwright. Clinical reasoning: survey of teaching methods, integration, and assessment in entry-level physical therapist academic education. *Physical Therapy*, 97(2):175–186, 2017.

[10] Karen Huhn, Sarah J Gilliland, Lisa L Black, Susan F Wainwright, and Nicole Christensen. Clinical reasoning in physical therapy: a concept analysis. *Physical Therapy*, 99(4):440–456, 2019.

[11] Michelle Daniel, Joseph Rencic, Steven J Durning, Eric Holmboe, Sally A Santen, Valerie Lang, Temple Ratcliffe, David Gordon, Brian Heist, Stuart Lubarsky, et al. Clinical reasoning assessment methods: a scoping review and practical guidance. *Academic Medicine*, 94(6): 902–912, 2019.

[12] Ana María de Santiago Nocito and Alberto García Lledó. El script concordance test como herramienta evaluativa. *RIECS: Revista de Investigación y Educación en Ciencias de la Salud*, 5(2):122–125, 2020.

[13] Lindsey Kojich, Stephanie A Miller, Katelyn Axman, Timothy Eacret, J Atticus Koontz, and Caroline Smith. Evaluating clinical reasoning in first year DPT students using a script concordance test. *BMC Medical Education*, 24(1):1–9, 2024.

[14] Nicotine M Otterman, MJ Maas, Sven K Schiemanck, Philip J Van Der Wees, and Gert Kwakkel. Development and validity of an innovative test to assess guideline-consistent clinical reasoning by physical therapists in stroke rehabilitation. *Journal of Rehabilitation Medicine*, 51(6):418–425, 2019.

[15] IFOMPT. IFOMPT in a nutshell, 2020. Disponible sur: https://www.ifompt.org/About+IFOMPT.html.

[16] Valérie Dory, Robert Gagnon, Dominique Vanpee, and Bernard Charlin. How to construct and implement script concordance tests: insights from a systematic review. *Medical Education*, 46(6):552–563, 2012.

[17] Stuart Lubarsky, Valérie Dory, Paul Duggan, Robert Gagnon, and Bernard Charlin. Script concordance testing: from theory to practice: AMEE guide No. 75. *Medical Teacher*, 35(3):184–193, 2013.

[18] Robert Gagnon, Bernard Charlin, Michel Coletti, Evelyne Sauvé, and Cees Van der Vleuten. Assessment in the context of uncertainty: how many members are needed on the panel of reference of a script concordance test? *Medical Education*, 39(3):284–291, 2005.

[19] Bernard Charlin, Robert Gagnon, Louis Sibert, and Cees Van der Vleuten. Le test de concordance de script, un instrument d'évaluation du raisonnement clinique. *Pédagogie Médicale*, 3(3): 135–144, 2002.

[20] Suzette Cooke, Jean-Francois Lemay, and Tanya Beran. Evolutions in clinical reasoning assessment: the evolving script concordance test. *Medical Teacher*, 39(8):828–835, 2017.

[21] Marie-France Deschênes, Bernard Charlin, Robert Gagnon, and Johanne Goudreau. Use of a script concordance test to assess development of clinical reasoning in nursing students. *Journal of Nursing Education*, 50(7):381–387, 2011.

[22] Caroline Faucher, Marie-Pier Dufour-Guindon, Gabrielle Lapointe, Robert Gagnon, and Bernard Charlin. Assessing clinical reasoning in optometry using the script concordance test. *Clinical and Experimental Optometry*, 99(3):280–286, 2016.

[23] Carole Lambert, Robert Gagnon, David Nguyen, and Bernard Charlin. The script concordance test in radiation oncology: validation study of a new tool to assess clinical reasoning. *Radiation Oncology*, 4(1):1–6, 2009.

[24] Eric Steinberg, Ethan Cowan, Michelle P Lin, Anthony Sielicki, and Steven Warrington. Assessment of emergency medicine residents' clinical reasoning: validation of a Script Concordance Test. *Western Journal of Emergency Medicine*, 21(4):978–984, 2020.

[25] Julie Subra, Bruno Chicoulaa, André Stillmunkès, Pierre Mesthé, Stéphane Oustric, and Marie-Eve Rougé Bugat. Reliability and validity of the script concordance test for postgraduate students of general practice. *European Journal of General Practice*, 23(1):209–214, 2017.

[26] Sibylle Vital, Claudine Wulfman, Félix Girard, Faleh Tamimi, Bernard Charlin, and Maxime Ducret. Script concordance tests: A call for action in dental education. *European Journal of Dental Education*, 25(4):705–710, 2021.

[27] Nicolas Fernandez, Amélie Foucault, Serge Dubé, Diane Robert, Chantal Lafond, Anne-Marie Vincent, Jeannine Kassis, Driss Kazitani, and Bernard Charlin. Learning-by-Concordance (LbC): introducing undergraduate students to the complexity and uncertainty of clinical practice. *Canadian Medical Education Journal*, 7(2):e104–e103, 2016.

[28] Amélie Foucault, Serge Dubé, Nicolas Fernandez, Robert Gagnon, and Bernard Charlin. Learning medical professionalism with the online concordance-of-judgment learning tool (CJLT): A pilot study. *Medical Teacher*, 37(10):955–960, 2015.

[29] Tsuen-Chiuan Tsai. Twelve tips for the construction of ethical dilemma case-based assessment. *Medical Teacher*, 39(4):341–346, 2017.

[30] Mohsen Tavakol and Reg Dennick. Making sense of Cronbach's alpha. *International Journal of Medical Education*, 2:53–55, 2011.