

Beyond the Basics of Clinical Outcomes Assessment: Selecting Appropriate Patient-Rated Outcomes Instruments for Patient Care

Alison R. Valier, PhD, ATC; Kenneth C. Lam, ScD, ATC
Interdisciplinary Health Sciences, A.T. Still University, Mesa, AZ

The fifth edition of the *Athletic Training Education Competencies* emphasizes the concepts of clinical outcomes assessment. In athletic training, clinical outcomes assessment, especially as it relates to patient-rated outcomes (PRO) instruments, is new, which produces uncertainty with regard to how to integrate PROs into athletic training education. Our goal was to review the concepts associated with selecting PRO instruments and to provide a teaching strategy for implementing these concepts into education programs. When selecting a PRO instrument, clinicians should follow a systematic process that evaluates a variety of criteria related to the development and performance of the instrument as well as the clinical utility of the instrument. Considering the importance of the selection process, athletic training educators may be unsure of strategies to guide their students through the process. The process of selecting PRO instruments is not unlike that used to select other clinical tools, such as tools with which to measure range of motion or strength. Selecting PRO instruments requires consideration of both essential elements (ie, instrument development, reliability, validity, responsiveness and interpretability, and precision) and clinical utility components (ie, acceptability, feasibility, and appropriateness). This manuscript provides key considerations for the selection criteria as well as a case scenario assignment that is used to demonstrate how the criteria can be applied to a clinical case. Creating assignment strategies to apply the concepts of clinical outcomes assessment into clinical practice may result in clinicians who appreciate the value of patient outcomes as integral components of patient care. The evaluation of the essential elements and clinical utility of an instrument provides a framework by which to select an appropriate PRO instrument. By using a case scenario assignment, athletic training educators can guide their students through the selection process and highlight important considerations when comparing PRO instruments.

Key Words: Patient-oriented evidence, quality of life, patient-centered care, education strategies, athletic training

The fifth edition of the *Athletic Training Education Competencies* emphasizes the concepts of clinical outcomes assessment.¹ Assessment of clinical outcomes is not new, and other health care professions have included these instruments as a component of patient care for years. In athletic training, clinical outcomes assessment, especially as it relates to patient-oriented outcomes, is new, and many athletic trainers may be unsure of the methods by which to teach and integrate these instruments into athletic training education and clinical practice.²⁻⁴

One of the central elements of clinical outcomes assessment is the ability to select an appropriate patient-rated outcome (PRO) instrument for patient care. Given the large number of available instruments, selecting an appropriate PRO instrument is essential for obtaining meaningful information to manage a patient and make valuable clinical decisions. When selecting a PRO instrument, clinicians should follow a systematic process that evaluates a variety of criteria, such as instrument

reliability, appropriateness, and friendliness, related to the development and performance of the instrument as well as the clinical utility of the instrument.⁵ Considering the importance of the selection process, athletic training educators may be unsure of strategies with which to effectively guide their students through the process. Therefore, the purpose of this manuscript is to review the concepts associated with selecting PRO instruments. A *selecting PRO instruments assignment* will be used to describe an assignment template and strategy for incorporating this material into athletic training programs.

INTEGRATING SELECTION CRITERIA INTO CLASSROOM LEARNING

The process of selecting PRO instruments is similar to those steps necessary to selecting clinical tools to measure range of motion, strength, swelling, or diagnostic tests based on their accuracy. Questions about the instrument are generally classified into two groups, with one group related to the

Dr Valier is currently an Associate Professor at A.T. Still University. Please address all correspondence to Alison R. Valier, PhD, ATC, A.T. Still University, Interdisciplinary Health Sciences, 5850 East Still Circle, Mesa, AZ 85206. arsnnyder@atsu.edu.

Full Citation:

Valier AR, Lam KC. Beyond the basics of clinical outcomes assessment: selecting appropriate patient-rated outcomes instruments for patient care. *Athl Train Educ J*. 2015;10(1):91–100.

Table 1. Criteria for Selecting Patient-Rated Outcomes (PRO) Instruments

Classification	Component
Essential elements	Instrument development Reliability Validity Responsiveness and interpretability Precision
Clinical utility	Acceptability Feasibility Appropriateness

essential elements of the instrument and with the second group of questions related to the *clinical utility* of the instrument.^{5,6} Essential elements of an instrument may address questions, such as whether the instrument was developed with sound theory, whether it is reliable and valid, whether it measures change over time in a way that matters to patients, or whether the instrument has adequate score precision.⁵ Questions related to the clinical utility of the instrument may include whether the instrument is appropriate for the intended purpose, whether it is easy to interpret and apply, whether it is liked by patients and clinicians, or whether the cost and time associated with the instrument are manageable.⁵ Both the essential elements and the clinical utility questions translate into the selection criteria that should be used to evaluate any measurement tool, including PRO instruments (Table 1). For the most part, these criteria are familiar to students because similar concepts are often incorporated into other courses, such as statistics, or discussions of clinician-based measures (eg, goniometer or dynamometer). However, application of the selection criteria as they relate to PRO measurement can be daunting, especially for those unfamiliar with them.

When teaching PRO instrument selection, there is a lot of information about each selection criteria that could be shared with students. The volume of information could be overwhelming. To begin, it may help to focus attention on those aspects of each selection criteria that specifically relate to PRO instruments. A focus on PRO-related aspects of each selection criteria will make teaching the information more straightforward and manageable. The greater challenge, though, is providing a meaningful experience through which students can take the basic concepts of the selection criteria and apply them to a clinically meaningful situation. An assignment that encourages integration is helpful to verify that the concepts have been learned and to provide evidence of the student's ability to use these concepts in practice. The remaining portions of this manuscript describe (1) each PRO instrument selection criteria, divided into essential elements and clinical utility, and (2) a case scenario assignment that can be used to apply the selection criteria to a clinical situation.

ESSENTIAL ELEMENTS FOR INSTRUMENT SELECTION

Instrument Development

The first element to evaluate when reviewing a PRO is instrument development. Developing an instrument is a complex process and should be conducted in a systematic, comprehensive, and logical manner. Disablement models have been used as a framework for designing instrument content,⁷

although this method is not the standard. While there are many methods to use to develop an instrument, key components include item generation and reduction, instrument/item testing and further reduction, and establishment of psychometric properties (eg, reliability and validity).⁸ These steps ensure that questions are developed and critically reviewed, so that the final instrument includes content that is valuable to the target patient population. Instrument development should also include input from clinician content experts and patients,⁹ and the instrument should be tested in both healthy and injured populations. A literature search for articles related to the development of the instrument is necessary when determining whether it was developed through a systematic and logical process.

Reliability

The second element to review for instrument selection is the reliability of the instrument. Reliability indicates the ability of an instrument to differentiate between people,⁹ such as different patient populations, and is necessary for measuring change over time, a key purpose of PRO instruments. When teaching students about instrument reliability, it is important to stress that reliability is not an inherent characteristic of an instrument and should be considered for each new population. For example, an instrument that is reliable in middle-aged adults may or may not be reliable in high school students. Additional reliability terms that should be introduced to students are *stability* and *internal consistency*.

Stability indicates the ability of the PRO to reproduce a score on repeated assessments at different times while the patient's health status remains unchanged.⁹ A stable PRO instrument is important so interpretation of scores accurately reflects current health status as opposed to error that is inherent in the PRO instrument. The Pearson product moment correlation coefficient (r) and the intraclass correlation coefficient (ICC) are 2 common measures of stability.⁹ Internal consistency is important when considering the reliability of PRO instruments because many instruments consist of more than 1 question to measure a single dimension of health (eg, physical or social).⁹ All questions related to a health domain should be homogeneous.

When evaluating internal consistency, a statistical technique to consider is Cronbach α , which represents the average correlation of all of the questions in the instrument.⁹ Higher Cronbach α values generally indicate higher internal consistency.¹⁰ An example of internal consistency, as it relates to outcomes instruments, is to consider the 9 International Knee Documentation Committee Subjective Knee Form (IKDC) questions related to function (items 9a–9i).⁸ Better internal consistency for function means that the function questions score similarly. Therefore, if an individual was experiencing difficulty with function, questions related to function should be scored consistently as more difficult. For example, if an individual has trouble with going up or down stairs, it is likely that the individual will also have difficulty with other functional tasks, such as running straight ahead, jumping and landing on the involved leg, and stopping and starting quickly. However, too-perfectly correlated questions (Cronbach $\alpha = 0.90$ – 1.0) produce a narrow measure of a domain, suggesting that the questions are redundant and unnecessary. For example, while the IKDC offers several tasks with varying

levels of difficulty within the physical functioning domain, an instrument that only includes a narrow range of functional tasks (eg, “Can you walk 5 steps?,” “Can you walk 10 steps?,” “Can you walk 1 block?”) would limit the clinician’s ability to comprehensively evaluate the physical function of the patient.

Students should be taught how to interpret reliability coefficients, such as the Pearson r , ICC, and Cronbach α . According to Portney and Watkins,¹⁰ poor, moderate, and good reliability are represented by values below 0.50, between 0.50 and 0.75, and above 0.75, respectively.¹⁰ Lower reliability values (eg, 0.7) may be appropriate for group-level data, and higher values (eg, 0.9) may be appropriate when looking at data for individuals.¹⁰ Typically, a higher reliability value is expected with individual data because the variability within an individual should be less than that between individuals (group data). Ultimately, the acceptable reliability values for an instrument are based on researcher or clinician judgment and should be selected to allow meaningful use.¹⁰

Validity

The validity of an instrument should also be considered before implementing a PRO instrument, because it is only with valid instruments that we can draw meaningful inferences from the reported scores.⁹ To validate a PRO instrument, evidence must show that the instrument evaluates the intended constructs. Like reliability, validity is not an inherent characteristic of an instrument, so validity evidence must be found for the target population. Three types of validity are often highlighted: content, criterion, and construct.¹¹ *Content validity* indicates the extent that a domain of interest (eg, physical function) is comprehensively sampled,¹¹ and evidence of comprehensive sampling frequently comes from expert and patient opinion. *Criterion validity* compares the PRO of interest to a “gold standard” instrument.⁵ Because there are no gold standard PRO instruments, these comparisons often occur when a shorter PRO instrument is made from a longer PRO instrument, with the longer instrument used as the gold standard. Quality Metric’s short form¹² instruments, which are widely used generic measures of health-related quality of life (HRQOL), provide an example of using a longer outcomes instrument (SF-36) as a gold standard for shorter instruments (eg, SF-12 and SF-8). While there is no true PRO gold standard, the SF-36 is often considered a gold standard of patient-reported HRQOL¹³ because it has been used extensively in medical research and tested in a variety of patient populations. Finally, *construct validity* is the evaluation of a theory that explains the association among behaviors and attitudes being studied.⁹ In relation to PROs, construct validity is used to confirm that unobservable factors (ie, constructs) of interest, such as pain or motivation, are captured by the instrument. Validity testing is an ongoing, unlimited process, and it is not uncommon to see a number of different validation studies related to a single PRO instrument.

Responsiveness and Interpretability

Evaluation of instrument responsiveness and interpretability is an important component of any PRO review. *Responsiveness* refers to the ability of an instrument to detect true change over time and is essential when administering PROs.^{5,11} In general, responsiveness can be considered from a statistical perspective or a clinical meaningfulness perspective. While

both perspectives are important, clinical meaningfulness is usually highlighted because it is useful for driving treatment decisions and may be easier to interpret than are more statistically based values (eg, standardized response mean, effect size). *Minimal clinically important difference* (MCID), also called *minimal important difference* or *clinically important difference*, is the smallest amount of change a patient perceives as beneficial and is a measure of responsiveness.¹⁴ Minimal clinically important difference values are typically presented as point values. For example, the Lower Extremity Functional Scale (LEFS) has a reported MCID value of 9 LEFS points.⁷ This value would tell you that to be comfortable that your patient experienced change that is likely perceived as beneficial, a change in LEFS scores of at least 9 points would need to be observed between the 2 time points. Essentially, the MCID provides a threshold for determining whether the patient perceives meaningful change from one time point (eg, initial appointment) to another (eg, 2-week follow-up appointment or discharge). Responsiveness values, such as MCID, are not inherent to the instrument, or static, and can change as a result of a variety of factors related to the methodological approach used to derive the MCID (eg, population studied).¹⁵ Therefore, the methods used to derive the MCID should be considered and a value should be used that was generated from a population as close to the one of interest as possible. One challenge in athlete health care is that there are not as many reports of MCID values associated with PROs in this population as in others, which affects the interpretability of these instruments when they are used with athletes.

In addition to MCID, other values provide information useful in interpreting PRO scores. For example, the *standard error of the measure* is a value that represents the error associated with a patient completing a PRO on one occasion.¹⁶ The *minimal detectable change* (MDC) is the value that represents the error associated with a patient completing a PRO on more than one occasion (eg, initial visit and discharge visit).¹⁶ Standard error of the measure and MDC, like MCID, are also both indicated typically by instrument point values. For example, the MDC for the LEFS is 9 LEFS points.⁷ Therefore, to be confident that the change in LEFS score reported from one time point (eg, initial visit) to another (eg, discharge) represented true change and not error, a difference of 9 LEFS points or more would need to be observed. In our examples using the LEFS, the MCID and MDC values were the same (ie, 9 LEFS points), but this is not always the case, and frequently these values will be different. While error values, such as the standard error of the measure and MDC, are statistical in nature and do not reflect clinically meaningful change, they do provide a level of confidence in interpreting scores, allowing clinicians to differentiate between true change in a patient’s health and error within the measure.

One point of caution when evaluating the responsiveness of an instrument is that there may be ceiling effects reported for some PRO instruments, which can impact the validity of the instrument. Ceiling effects occur when a percentage of the population scores at the highest level of health, even when suffering from a health condition, such as an injury.⁵ Athletes tend to function at higher levels of health. Thus, there is concern that athletes will rate their health high on a PRO when injured, and the instrument will be unable to capture improvement in health status on the instrument because the

top score was already achieved. The issue of ceiling effects in PRO instruments as it relates to athletes warrants further exploration.

Precision

The final element to consider when selecting PRO instruments is the precision of the response categories. Common response options are binary, visual analogue, adjectival, and Likert scales.⁹ Users of PRO instruments should find a balance between the simplicity of the response options and the quality of information received. The simplest scale is binary and offers 2 response options, such as *yes* or *no*. Binary scales are quick and easy to complete but are limited because they provide little depth about the information gathered. Visual analogue scales are simple and are constructed by using a line of fixed length (eg, 10 cm) that is anchored on the ends with extreme value options (eg, *no pain* and *extreme pain*). In contrast, more complex answer systems, such as adjectival and Likert scales, may require more thought to answer but provide deeper information about the questions and topics of interest. Adjectival and Likert scales are similar in that they contain descriptors along a continuum for patients to select. The one difference is that adjectival scales are unipolar (eg, ranging from a minimum to a maximum amount) and Likert scales are bipolar (eg, ranging from *strongly agree* to *strongly disagree*).⁹ With more complex scales, the preferred number of response options should be considered. While there is no maximal number of response options, responders may not be able to differentiate the responses on scales with more than 7 options.⁹

CLINICAL UTILITY CONSIDERATIONS FOR INSTRUMENT SELECTION

In addition to the essential elements of an instrument, the overall clinical utility of the PRO instrument must be considered and evaluated. For clinical utility, review should determine whether the instrument can be easily incorporated into routine patient care without hindering the natural workflow of daily clinical practice and whether the instrument supports the specific health care goals of the patient, clinician, and clinic.^{5,6} While essential measurement properties can be assessed through development papers and literature searches, the clinical utility of a PRO instrument is primarily based on the clinician's best clinical judgment after considering the unique characteristics of the patient, patient population, and clinical setting.

Acceptability

An element for determining the clinical utility of a PRO instrument is assessing the overall acceptability, or patient friendliness, of the instrument. The clinician should choose an instrument that will minimize patient burden, maximize response rates, and obtain quality information from the patient.⁵ Patient friendliness is often characterized by the total time it takes a patient to complete the PRO instrument and is influenced by the total number of items in the measure and the interpretability of the items by the patient. For example, a PRO instrument like the Pediatric Outcomes Data Collection Instrument, which is a generic, pediatric-specific measure of 83 items, may have low acceptability in a busy athletic training clinic because of the time it would take the patient to

complete. Similarly, if items are difficult for a patient to interpret, frustration or confusion may result in the patient who needs more time to complete the PRO. Furthermore, it may be helpful to evaluate a PRO for readability to ensure that most people understand the terms used within the instrument. The readability of PROs can be evaluated through informal or formal methods. Informally, readability can be evaluated by reviewing each question on the instrument and making a personal judgment about how easy the question is to understand. A formal approach requires the use of computer software to produce readability scores. For example, the Flesch reading ease score (higher percentages indicate better readability; 75% = plain English) and the Flesch-Kincaid readability index (eighth grade or below is a common threshold for an average American) can be calculated using Microsoft Word. Issues with long length and readability may hinder a clinician's ability to maximize overall response rate (eg, patient does not complete the instrument) or attain quality information from the patient (eg, patient completes the PRO but loses interest in answering the questions toward the end).

While an 83-item instrument may be too long for use in an average athletic training clinic, a 5-item or 10-item instrument is not necessarily a better alternative. The primary goal of incorporating PRO instruments into patient care is to gather useful information from the patient's point of view.^{2,4} Thus, fewer items in an instrument equate to less information attained from the patient. For example, consider a comparison between the Global Rating of Function and the function-specific section of the IKDC (items 9a–9i). The Global Rating of Function is a single-item question that asks the patient to rate the amount of use of an injured body part on a scale from 0 (*no use*) to 100 (*full use*); the IKDC has a specific function section that includes several questions related to different types of lower extremity function (eg, ability to go up and down stairs, squat, jump, and land). While both instruments aim to characterize overall function of the patient, the multi-item IKDC provides more specific, meaningful information related to function for patient care decisions and development of patient goals than does the single-item Global Rating of Function. Thus, an appropriate balance between the total number of items and the depth of information obtained from the patient must be considered when evaluating the overall acceptability of a PRO instrument. That balance depends on the judgment of the clinician.

One important point is that overall acceptability of a PRO instrument will be influenced by the specific characteristics of individual patients and patient populations. For example, the acceptability of a PRO instrument will differ for a 15-year-old high school athlete and a 35-year-old recreational athlete as a result of reading comprehension levels and maturity differences. Furthermore, questions in a PRO instrument may make the patient uncomfortable, depending on participant age and maturity level. For example, one question in the Disability of the Arm, Shoulder and Hand questionnaire asks how the injury has impacted sexual activities, which may be inappropriate for a younger patient population. While a concern, these types of questions are relatively uncommon and would not automatically disqualify an instrument from use in patient care. Often, PRO instruments allow for score calculations with missing items, so the clinician would still be able to accurately calculate the overall instrument score even if an

inappropriate item was omitted. Because of the characteristics of individual patient populations, each item should be reviewed and evaluated to determine the acceptability of a PRO instrument for its intended population and patient care use.

Feasibility

Another clinical utility component that warrants review is the feasibility or clinician friendliness associated with the PRO instrument.^{5,6} When assessing feasibility, attention should be focused on whether the instrument can be incorporated into practice without interrupting the natural workflow of the clinic or impeding the clinician's ability to provide optimal patient care. The primary goal is to identify an instrument that will allow the clinician to collect information from the patient without overburdening the clinician. In general, clinician friendliness is characterized by the overall ease of use of the instrument. When evaluating ease of use, several factors should be considered, such as required clinician training to properly use the PRO instrument, potential costs associated with the use of the instrument, the need for the clinician to supervise the patient during completion of the instrument, the time needed to score a completed instrument (eg, simple sum, average or percentages, reverse scoring, score transformations), and the interval between repeated administration of the instrument (eg, recall period of the instrument). Training clinicians to properly interpret scores on instruments with many subscales or composite scores, such as the SF-12, may be necessary. Additionally, more complex scoring systems for a PRO, such as PROs that result in multiple subscale scores like the SF-12, may require training to ensure that clinicians understand the actual scoring process. Usage fees, licensing agreements, and user agreements may be a requirement of use for some PROs. For example, the SF-12 and the Pediatric Quality of Life Instruments have licensing agreements and associated fees depending on the intended use of the instrument (eg, research versus clinical practice).

In general, there is no set rule on how often to administer a PRO, and time of administration may depend on the severity of the condition as well as the instrument's recall period. For example, a patient with an anterior cruciate ligament (ACL) tear may initially complete a PRO at routine therapy visits about once every 1 to 2 weeks. As the patient progresses (eg, after 1 month of treatment), administrations may be less frequent, such as 1 time per month until discharge. A patient with a less severe injury, such as an ankle sprain with an expected recovery of 2 weeks, may need fewer PRO assessments. For example, it may be that a clinician administers the PRO at the initial and final appointments only.

Appropriateness

The last element for clinical utility is appropriateness. *Appropriateness* of a PRO addresses whether the instrument would support the general and specific goals of patient care. Evaluating whether an instrument is appropriate for its intended purpose can be driven by many factors, including whether the target patient population aligns with the patient population used for instrument development, the variables or content areas of interest to the patient case (eg, disablement levels, HRQOL domains), and the global purpose for utilizing

the PRO (eg, characterizing quality of care or optimizing patient-centered care).^{5,6} The dimensions of disablement allow the clinician to view the patient from impairment, function, and disability perspectives, whereas HRQOL allows the clinician to view the patient based on different health domains, such as physical, psychological, and social functioning domains. These different factors may lead a clinician to choose one instrument over another.

Measurement properties established in one patient population cannot be transferred or assumed to be the same for another patient population. For example, an instrument that has been found valid and reliable in one patient population (eg, adolescent athletes) is not necessarily valid and reliable in a different patient population (eg, adult athletes). In fact, many PRO instruments have different versions of the same instrument to account for patient population differences. Recently, adolescent versions of the IKDC¹⁷ and Knee Injury and Osteoarthritis Outcome Score (KOOS)¹⁸ were developed to address the increasing number of adolescent knee injuries. Thus, when selecting a PRO instrument, measurement properties should be verified for specific target populations, and the instrument should have demonstrated usefulness in the target population.

A primary aim of utilizing a PRO instrument is to support whole-person health care, which can be framed under the concepts of disablement and HRQOL.^{2,4} To determine which instrument is most appropriate, each question within an instrument should be reviewed by the clinician and classified within a disablement dimension or an HRQOL domain. Depending on the types of changes the clinician intends to capture over the course of care, one instrument may be more appropriate for patient care and patient care goals than another.

Finally, appropriateness can also be viewed in terms of the global purpose of use. From a global perspective, the primary goals of clinical outcomes assessment are to help clinicians characterize the quality of their patient care (ie, the effectiveness) and to optimize their ability to provide patient-centered care. While the incorporation of PRO instruments allows the clinician to achieve both of these goals, emphasis will likely be on one of them. For example, if establishing the effectiveness of a particular intervention (eg, manual therapy after knee injury) is the primary goal, an instrument should be selected that applies to a wide range of patients, so that greater numbers of patients could be included in the assessment (larger "n"). Using an instrument that has broad applicability to patients and requiring that clinicians use that one instrument with their patients may result in determining the effectiveness of an intervention faster because of the larger number of patients completing the same PRO, the data for whom could be aggregated. In contrast, if patients with knee injuries completed a variety of PRO instruments (eg, one-third of the patients completed the IKDC, KOOS, or Lysholm) instead of one instrument (eg, the IKDC), that data could not be easily aggregated because the instruments ask different questions and produce scores that are only relevant to that instrument. Therefore, the use of multiple instruments across participants for the purpose of making the determination of treatment effectiveness would be a longer and more inefficient process than would be selecting and using one instrument that could be applied to many patients. If the goal

is to optimize patient-centered care, an instrument should be selected that most closely matches the specific deficits exhibited by the patient or that addresses content most valued or important to that patient. For optimizing patient-centered care, the utilization of several different knee-specific instruments across several patients with different and distinct limitations would be appropriate and beneficial.

Case Scenario Assignment

A strategy for teaching students to select PRO instruments is to create an assignment that applies the concepts of selecting instruments to a clinical scenario. The assignment can be structured in many different formats, but a simple approach is to ask students to select 2 specific PRO instruments that apply to a real or fictitious, but realistic, clinical scenario for critique, comparison, and presentation to the class. The remaining sections of this paper provide an example scenario that could be used to complete the assignment.

The following question shapes this clinical scenario: In athletes with recent ACL sprains, is the KOOS¹⁹ or the IKDC⁸ better for measuring changes meaningful to patients?

To complete the assignment, students will need to search the literature for articles that address the instrument selection criteria. The predominance of information used to evaluate the instruments in the case scenario came from articles by Irrgang et al,⁸ Roos et al,¹⁹ and van Meer et al²⁰ and by reviewing the instruments in general.

Essential Elements. Instrument Development. To evaluate instrument development, students should find development articles related to their instruments of interest. For the case scenario, Irrgang et al⁸ had a discussion of the development of the IKDC, which included details about the purpose of the instrument, definition of constructs, item generation and pilot testing, item selection, and evaluation. Similarly, Roos et al¹⁹ reported on the development of the KOOS and highlighted the basic theory and purpose of the instrument, use of an expert panel, item generation and pilot testing, and evaluation, which all speak to a rigorous, detailed, and complete process. When evaluating the development of the IKDC and KOOS, both contain elements that suggest a rigorous, detailed, and complete instrument, and, as a result, we may conclude that development for these instruments is comparable. Therefore, instrument development is likely not the determining factor in instrument selection for this scenario. Unfortunately, a detailed instrument development process is not always completed, and students must be able to identify limitations in instruments that lack essential developmental elements, such as those that did not use patients in item generation or test the instrument in an injured patient population or those that have no information on instrument development.

Reliability. As was the case for instrument development, students should search the literature for development articles for their instruments, especially as they relate to psychometric properties. Additionally, articles should be as closely related to the patient population of interest as possible, although these can be hard to find when the target population is young, relatively healthy athletes. For example, in the case scenario, we are interested in athletes who have suffered a recent ACL

injury, and, thus, our focus should be on studies about the reliability of the IKDC and KOOS in ACL-injured athletes. van Meer et al²⁰ studied both the IKDC and KOOS in patients, aged 18 years and older, who had recent ACL ruptures and were seeking services from an orthopaedic surgeon. Both the IKDC (ICC = 0.93) and KOOS (ICC = 0.81–0.87 for pain, symptoms, activities of daily living, sports/recreation, and quality-of-life subscales) were determined to be reliable (ICC \geq 0.81). While our selected article is not the exact fit for our population of interest (ie, college athletes), it does provide us with some information that speaks to the reliability of the instruments. Comparing the reliability of the IKDC and KOOS using information from our selected article would lead us to conclude that both instruments are reliable, and, therefore, reliability is likely not the determining factor in instrument selection for this scenario.

Validity. Development articles will also be of value to students as they work to evaluate the validity of their chosen instruments in their population of interest. van Meer et al²⁰ measured both content and construct validity of the IKDC and KOOS instruments, so this article will be helpful for our case scenario. To evaluate content validity, the authors had experts and patients identify the number of questions relevant to patients with ACL ruptures on the IKDC total and KOOS subscale scores. The IKDC and KOOS scores were regarded as relevant when more than 75% of the raters endorsed the individual questions as relevant to patients with ACL ruptures.²⁰ Construct validity was measured by comparing the IKDC and KOOS to other validated instruments that were designed to measure similar factors and complaints, such as a visual analogue scale for pain, the SF-36 subscales, and the Lysholm instrument.²⁰ Hypotheses were generated to indicate the direction and magnitude of the correlation coefficients between the IKDC and KOOS instruments, where $\geq 75\%$ of confirmed hypotheses indicated good construct validity. Table 2 presents the content and construct validity, as reported by van Meer et al,²⁰ for the IKDC and KOOS. Criterion validity was not tested by van Meer et al,²⁰ since there are no gold standards for these instruments. Based on these validity results, the IKDC should be slightly favored over the KOOS because of the limited content and construct validity (ie, $< 75\%$ hypotheses matched) for some of the KOOS subscales.

Responsiveness and Interpretation. Students should seek articles related to psychometric properties when looking to evaluate responsiveness (eg, MCID) or error (standard error of the measure or MDC) of the instruments in their population of interest. Finding responsiveness and interpretability information may require students to locate additional articles, and for some PROs, no responsiveness information is available. Irrgang et al²¹ studied the responsiveness of the IKDC in patients with a variety of knee injuries (eg, ligamentous, meniscal) and conditions (eg, osteoarthritis) and reported an MCID value of 11.5 IKDC points in this population. Therefore, a change in IKDC scores from one time (eg, initial appointment) to another (eg, 2-week follow-up appointment) of more than 12 IKDC points would provide some confidence that the patient had perceived a meaningful change in health status. Similarly, the minimal perceptible clinical improvement, a value like the MCID, is about 10 points for the KOOS.²² Because both the IKDC and KOOS have reported meaningful change values, responsiveness is

Table 2. Essential Element Considerations for Instrument Selection

Element	IKDC	KOOS	Decision
Instrument development	Robust process used (eg, instrument purposed, defined constructs, pilot testing, reliability and validity testing) ⁸	Robust process used (eg, theory and purpose, expert panel, pilot testing, evaluation of measurement properties) ¹⁹	Both instruments appear to be developed with robust processes.
Reliability	ICC = 0.93 ²⁰	ICC > 0.81 ²⁰	Both instruments have acceptable reported reliability.
Validity			
Construct	>75% Hypotheses matched ²⁰	>75% Hypotheses matched for sport/recreational subscale only ²⁰	Both instruments were tested for construct and content validity, and, overall, both are valid instruments. The IKDC may be more desirable than the KOOS: more hypotheses matched (construct validity); more relevant items (content validity).
Content	>75% Relevant items ²⁰	>75% Relevant items for symptoms, sports/recreational, and quality of life subscales ²⁰	
Responsiveness and interpretability	MCID = 11.5 points ²¹	MPCI = 10 ²²	Both instruments have reported meaningful change scores.
Precision	Binary, adjectival, modified VAS	Adjectival	Both instruments use measurement systems with acceptable precision (predominantly adjectival).

Abbreviations: ICC, intraclass correlation coefficient; IKDC, International Knee Documentation Committee; KOOS, Knee Injury and Osteoarthritis Outcome Score; MCID, minimal clinically important difference; MPCI, minimal perceptible clinical improvement; VAS, visual analogue scale.

likely not the determining factor in instrument selection for this scenario.

Precision. Evaluation of precision is less about the literature and more about a qualitative review of the PRO questions and response options. Students should focus on identifying the response options and be able to discuss the benefits or weaknesses of the styles incorporated in their instruments of interest. The IKDC uses binary, adjectival, and modified visual analogue scales as response options, whereas the KOOS uses only adjectival scales. Although the IKDC uses a couple of different answer styles for the questions, the predominance of questions in both instruments is adjectival. Therefore, precision is likely not the determining factor for instrument selection in this scenario.

Table 2 highlights the essential considerations for instrument selection, with a general analysis of each element for the class assignment.

Clinical Utility. Acceptability. Evaluating acceptability of a PRO requires a qualitative assessment of the instrument. To start, the student should evaluate the number of questions in the PRO. For the case scenario, the IKDC is composed of 18 items and the KOOS is composed of 42 items.^{8,19,23} As a result, the length of the IKDC may be preferred over the

KOOS because the IKDC will probably require less completion time, which equates to less burden on the patient. Patient-rated outcome instruments should also be evaluated for readability using informal and/or formal methods. In the case of the IKDC and KOOS, an informal review indicates that there are no red flags for readability because the instruments are both fairly easy to understand. Evaluation of each question for content indicates that there are no questions that may make the patient feel uncomfortable. Based on these findings, the IKDC seems to have better acceptability than the KOOS because it only has 18 questions and offers an appropriate balance between the total number of items and the ability to capture the appropriate amount of information for informing patient care decisions and developing patient care goals. Additionally, the questions on the IKDC are easy to read and should be perceived as comfortable to answer by patients.

Feasibility. The clinician friendliness of PRO instruments can also be reviewed through a review of the key considerations related to the barriers to implementation. In the case scenario, neither the KOOS nor the IKDC requires formal training to administer, and licensing fees are not required for their use.^{8,19,23} Second, the role of the clinician appears to be minimal for both instruments because neither requires the clinician to supervise the patient during completion or to

Table 3. Clinical Utility Considerations for Instrument Selection

Clinical Utility	IKDC	KOOS	Decision
Acceptability			
No.	18 items	42 items	The IKDC has fewer items and takes less time to complete, making it more acceptable than the KOOS.
Time to complete	5–7 min	10 min	
Readability ^a	Sixth grade	Fourth grade	
Comfort level issues	None	None	
Feasibility			
Ease of use	No training or supervision required; easy to explain	No training or supervision required; easy to explain	Few differences between the 2 instruments in terms of feasibility.
Role of clinician	No questions for clinician to complete. Recall period: current and past 4 wk	No questions for clinician to complete. Recall period: 1 wk	Both instruments are likely feasible for the case scenario.
Time to score	5 min by hand	5 min using associated software	Both instruments are equally quick to score. No difference in cost.
Associated costs	None; free to use	None; free to use	
Appropriateness			
Intended patient populations	Variety of knee injuries (eg, ligament, meniscal, articular cartilage, patellofemoral pain)	Knee osteoarthritis and knee injuries that may lead to osteoarthritis (eg, anterior cruciate ligament, meniscal, and chondral injuries)	Little difference between the 2 instruments in terms of appropriateness.
Demonstrated use for other patient populations	Ligament injuries and reconstruction, meniscal tears and surgery, cartilage lesions, osteochondritis dissecans, traumatic knee dislocation, platelet-rich plasma injections, and lateral release	Ligament injuries and reconstruction, meniscal tears and surgery, cartilage lesions, osteochondritis dissecans, total knee replacements, pharmacologic therapy, and glucosamine supplementation	Both instruments are likely appropriate for the identified patient case scenario.
Disablement levels	Impairment: 4 items Function: 13 items Disability: 1 item	Impairment: 19 items Function: 22 items Disability: 1 item	Both instruments emphasize function, with the KOOS also emphasizing impairment. Fewer questions on the IKDC may make it more desirable than the potentially more comprehensive KOOS.
Health-related quality-of-life dimensions	Physical: 17 items Psychological: 0 items Social: 1 item	Physical: 37 items Psychological: 2 items Social: 3 item	
Global purpose of use	Appropriate for characterizing quality of care and providing patient-centered care	Appropriate for characterizing quality of care and providing patient-centered care	Both instruments seem to capture the physical dimension of HRQOL. Both instruments would support global patient care goals.

Abbreviations: HRQOL, health-related quality of life; IKDC, International Knee Documentation Committee; KOOS, Knee Injury and Osteoarthritis Outcome Score.

^a Based on Flesch-Kincaid Readability (Microsoft Word).

complete a portion of the instrument.^{8,19,23} The time to score the IKDC and the KOOS seems comparable; however, for the KOOS, scoring software is available, which, if used, may decrease the burden on the clinician (eg, entry of question

scores into the computer).^{8,19,23} The most visible difference between the 2 instruments is that the IKDC asks the patient to recall health status over the past 4 weeks, whereas the KOOS has patients recall health status over the past week.

Differences in recall period would probably not influence the decision for selecting a PRO instrument for ACL patients because serial measures would likely be taken about a month apart. Thus, both the IKDC and KOOS appear to be suitable instruments, and feasibility does not appear to be a determining factor for instrument selection for this case scenario.

Appropriateness. Students should evaluate the appropriateness of their PRO instruments through a targeted literature search. For example, in our case scenario development articles indicate that the IKDC is intended for a wide variety of knee injuries, while the KOOS is primarily intended for patients with knee osteoarthritis and injuries that may lead to knee osteoarthritis (eg, ACL injury).^{8,19} Since their development, both instruments have been shown to be useful in other patient populations (Table 3). Finding that an instrument created for one purpose (eg, knee osteoarthritis) is applicable for other purposes (eg, ACL-injured patients) is not uncommon. Both the IKDC and KOOS appear appropriate for people who have suffered an ACL injury, and, therefore, appropriateness of population does not appear to be a determining factor for instrument selection for this case scenario.

Students should also evaluate the IKDC and KOOS according to disablement model framework and HRQOL domains. Both the IKDC and the KOOS emphasize the impairments and functional limitations of disablement levels and the physical functioning of HRQOL. The KOOS does seem to capture more information about psychological and social health domains than the IKDC because of the number of questions used to measure these areas. However, if the purpose of using the PRO is to measure changes in disablement or HRQOL, using a generic PRO instrument in conjunction with a specific instrument would be more practical because generic instruments tend to target global concepts of health, such as HRQOL. While adding 2 instruments may add the burden of time to the clinician and patient, the types of information received from the generic and specific PROs are complementary. Additionally, few specific instruments target multiple dimensions of disablement or quality of life; thus, it may be beneficial to include both a generic and a specific PRO in patient care. Overall, the IKDC and KOOS capture similar information regarding disablement and HRQOL, and, therefore, appropriateness in terms of a whole-person perspective does not appear to be a determining factor for instrument selection for this case scenario.

Finally, the PROs should be evaluated for the appropriateness to the intended purpose of use of the instruments. In our scenario, the goal is to evaluate changes in patients who have recently suffered an ACL injury and speaks more to the individual patient case. Patients with ACL injuries may have common but different injury presentations, so selecting a standard instrument for all may not be wise if the aim is to deliver patient-centered care that responds to the individual needs and preferences of the patient. However, selecting one standard instrument may be useful because it would allow aggregation of all outcomes information on patients with ACL and other knee injuries and would provide an opportunity to use the information for other purposes, such as for quality control or for evaluating treatment effectiveness. Both the IKDC and KOOS are appropriate for characterizing

quality of care and for providing patient-centered care, and, therefore, appropriateness in terms of global use does not appear to be a determining factor for instrument selection in this case scenario.

Table 3 highlights the clinical utility considerations for instrument selection and a general analysis of each selection element for the class assignment.

CONCLUSIONS

Once students have reviewed all of the essential and clinical utility elements for instrument selection, students should summarize their findings and identify their decision in either oral presentation or written format. While either format is educational, an oral presentation provides an interactive opportunity to engage the presenting student, classmates, and instructor in fine-tuning concepts and highlighting meaningful and subtle differences between instruments. An additional assignment option is to have students create a “key considerations” document that aggregates all of the students’ reviews of instruments into a single large document, such as an Excel file. This combined document could then be made available to students to use when considering instruments for future, real-world patient cases.

The increased use of PRO instruments in athlete health care necessitates the meaningful exposure of athletic training students to relevant content, such as the selection of outcomes instruments. Offering creative assignments that encourage students to apply the new material also helps to ensure concepts are understood and can be applied to patient care.

REFERENCES

1. National Athletic Trainers’ Association. *Athletic Training Education Competencies*. 5th ed. Dallas, TX: National Athletic Trainers’ Association; 2011.
2. Snyder AR, Parsons JT, Valovich McLeod TC, Bay RC, Michener LA, Sauers EL. Utilizing disablement models and clinical outcomes assessment to enable evidence-based athletic training practice: part I—disablement models. *J Athl Train*. 2008; 43(4):428–436.
3. Snyder AR, Valovich McLeod TC, Sauers EL. Defining, valuing, and teaching clinical outcomes assessment in professional and post-professional athletic training education programs. *Athl Train Educ J*. 2007;2(Apr–Jun):31–41.
4. Valovich McLeod TC, Snyder AR, Parsons JT, Bay RC, Michener LA, Sauers EL. Utilizing disablement models and clinical outcomes assessment to enable evidence-based athletic training practice: part II—clinical outcomes assessment. *J Athl Train*. 2008;43(4):437–445.
5. Fitzpatrick R, Davey C, Buxton MJ, Jones DR. Evaluating patient-based outcome measures for use in clinical trials. *Health Technol Assess*. 1998;1(14):1–69.
6. Snyder AR, Valovich McLeod TC. Selecting patient-based outcome measures. *Athl Ther Today*. 2007;12(6):12–15.
7. Binkley JM, Stratford PW, Lott SA, Riddle DL. The Lower Extremity Functional Scale (LEFS): scale development, measurement properties, and clinical application. North American Orthopaedic Rehabilitation Research Network. *Phys Ther*. 1999; 79(4):371–383.

8. Irrgang JJ, Anderson AF, Boland AL, et al. Development and validation of the International Knee Documentation Committee Subjective Knee Form. *Am J Sports Med.* 2001;29(5):600–613.
9. Streiner DL, Norman GR. *Health Measurement Scales: A Practical Guide to their Development and Use.* 3rd ed. New York, NY: Oxford University Press; 2003.
10. Portney LG, Watkins MP. *Foundations of Clinical Research: Applications to Practice.* 2nd ed. Upper Saddle River, NJ: Prentice Hall Health; 2000.
11. Suk M, Hanson BP, Norvell DC, Helfet DL. *Musculoskeletal Outcomes Measures and Instruments.* Vol 1. Dübendorf, Switzerland: AO Foundation Publishing; 2009.
12. Quality Metric. SF Health Surveys Web site. <http://www.qualitymetric.com/WhatWeDo/SFHealthSurveys/tabid/184/Default.aspx>. Accessed April 18, 2013.
13. Vickrey BG, Hays RD, Harooni R, Myers LW, Ellison GW. A health-related quality of life measure for multiple sclerosis. *Qual Life Res.* 1995;4(3):187–206.
14. Beaton DE, Bombardier C, Katz JN, et al. Looking for important change/differences in studies of responsiveness. OMERACT MCID Working Group. Outcome Measures in Rheumatology. Minimal Clinically Important Difference. *J Rheumatol.* 2001;28(2):400–405.
15. Michener LA, Snyder Valier AR, McClure PW. Defining substantial clinical benefit for patient-rated outcome tools for shoulder impingement syndrome. *Arch Phys Med Rehabil.* 2013; 94(4):725–730.
16. Michener LA, Leggin BG. A review of self-report scales for the assessment of functional limitation and disability of the shoulder. *J Hand Ther.* 2001;14(2):68–76.
17. Schmitt LC, Paterno MV, Huang S. Validity and internal consistency of the International Knee Documentation Committee Subjective Knee Evaluation Form in children and adolescents. *Am J Sports Med.* 2010;38(12):2443–2447.
18. Orqvist M, Roos EM, Brostrom EW, Janarv PM, Iversen MD. Development of the Knee Injury and Osteoarthritis Outcome Score for children (KOOS-Child): comprehensibility and content validity. *Acta Orthop.* 2012;83(6):666–673.
19. Roos EM, Roos HP, Lohmander LS, Ekdahl C, Beynon BD. Knee Injury and Osteoarthritis Outcome Score (KOOS)—development of a self-administered outcome measure. *J Orthop Sports Phys Ther.* 1998;28(2):88–96.
20. van Meer BL, Meuffels DE, Vissers MM, et al. Knee injury and osteoarthritis outcome score or International Knee Documentation Committee Subjective Knee Form: which questionnaire is most useful to monitor patients with an anterior cruciate ligament rupture in the short term? *Arthroscopy.* 2013;29(4): 701–715.
21. Irrgang JJ, Anderson AF, Boland AL, et al. Responsiveness of the International Knee Documentation Committee Subjective Knee Form. *Am J Sports Med.* 2006;34(10):1567–1573.
22. Roos EM, Lohmander LS. The Knee Injury and Osteoarthritis Outcome Score (KOOS): from joint injury to osteoarthritis. *Health Qual Life Outcomes.* 2003;1:64.
23. Collins NJ, Misra D, Felson DT, Crossley KM, Roos EM. Measures of knee function: International Knee Documentation Committee (IKDC) Subjective Knee Evaluation Form, Knee Injury and Osteoarthritis Outcome Score (KOOS), Knee Injury and Osteoarthritis Outcome Score Physical Function Short Form (KOOS-PS), Knee Outcome Survey Activities of Daily Living Scale (KOS-ADL), Lysholm Knee Scoring Scale, Oxford Knee Score (OKS), Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC), Activity Rating Scale (ARS), and Tegner Activity Score (TAS). *Arthritis Care Res (Hoboken).* 2011;63(suppl 11):S208–S228.