APTA Critical Appraisal Tool for Experimental Intervention Studies (CAT-EI)

Organization: This critical appraisal tool (CAT) was designed for guideline developers, researchers, clinicians, students and faculty to assist with assigning levels of evidence to the results of intervention studies. The CAT-EI is divided into 4 major sections.

- Part A includes the contextual information with regard to study purpose, location and IRB compliance.
- Part B, questions 1-12, assesses the overall quality of the research design.
- Part B, questions 13-20, assesses the individual outcome measures from the study, as selected by the reader.
- Part C provides prompts to assess the impact of the study

Uniqueness: The tool is unique from other available critical appraisal checklists. Most checklists evaluate the research design characteristics without separately considering the strength of the measures used. This CAT allows the reader to select specific outcome measures reported in an intervention study and separately evaluate the level of evidence afforded by those measures within the context of the study design. In this way, a single study with 3 outcome measures, one each with strong, moderate and weak psychometric properties, could potentially yield 3 different levels of evidence.

Rating Criteria: The criteria that should be used to rate each item are explained in a list preceding the tool and are aligned by question number. The criteria account for either or both the presence of information that should be reported and/or the quality of that information as it relates to potential bias that may affect the reader's confidence in the study's design and outcomes. It is assumed that the user is familiar with the terminology of research, statistics and evidence based practice methodologies. Some of the criteria provide terminology to consider, but the explanations are not full descriptions of the concepts; rather they are solely provided as rating criteria. Users are encouraged to look up any terminology that is unfamiliar.

Users: The CAT-EI was designed for both individuals and groups of literature appraisers, and for faculty who are teaching students and clinicians critical appraisal skills.

> **Individual Use**: When the CAT-EI is used by individual clinicians to rate the strength of evidence in one or more articles to make clinical decisions about the use of an intervention, the clinician will need to make their own determinations as to whether the study reports acceptable levels of potential bias, validity, reliability, and adherence to protocol.
>
> **Group Use:** When the CAT-EI is used by groups of appraisers for the creation of evidence-based clinical practice guidelines or systematic reviews, the group will need to agree on what constitutes the acceptable levels of potential bias, validity, reliability, and adherence to protocol. We recommend that each article is read by 2 or more appraisers who compare their ratings to achieve consensus, and that a 3rd reader or a group discussion ensue if consensus could not be reached.

Quality Ratings: The quality evaluations of the overall design and the individual outcomes are used to determine whether the study results are of high, acceptable, low or unacceptable quality. Definitions are adapted from Martin et al. (2014), however groups may choose other criteria as it fits their needs and should thus define them for publication purposes.

Tallies: The CAT-EI provides fields to tally the criteria ratings from the 2 Part B sections. Users have the option of using a numerical tally of the rated items to assign levels of evidence if their methodology requires that, but are advised that this tool equally weights all items and that may not be valid with regard to risk of bias. Instructions for accurate tallies are at the end of the CAT-EI forms.

Sandra L Kaplan PT, DPT, PhD Christine M. McDonough PT, PhD Matt Elrod PT, DPT, MEd, NCS

Martin RL, Davenport TE, Reischl SF, McPoil TG, Matheson JW, Wukich DK, McDonough, CM. Heel Pain-Plantar Fasciitis: Revision 2014. Clinical Practice Guidelines Linked to the International Classification of Functioning, Disability and Health from the Orthopaedic Section of the American Physical Therapy Association. J Orthop Sports Phys Ther. 2014; 44(11):A1-23.

Last Updated: 1/30/16
Contact: practice-dept@apta.org

APTA Critical Appraisal Tool for Experimental Intervention Studies

Please see below for instructions to complete this tool:

Instructions: A. Study Question and Design	
A.1	*General purpose: List the elements of the identified question using PICOT.*
Instructions: B. Quality Evaluation-Overall Design	
B.1	*Select "yes" if the inclusion criteria are described in sufficient detail.* *Select "no" if the description lacks clarity.*
B.2	*Select "yes" if the exclusion criteria are described in sufficient detail or if the author(s) state that there were no exclusion criteria.* *Select "no" if the description lacks clarity.*
B.3	*Select "yes" if recruitment bias is minimized to an acceptable level such that the resulting sample adequately represents the target population. Strategies include recruiting from multiple sites and advertising using multiple modes to ensure adequate representation of subject characteristics (e.g. ages and diagnostic severity).* *Select "no" if the level of potential bias significantly compromises your confidence in the results.*
B.4	*Select "yes" if an a priori power analysis is used to calculate the <u>minimum</u> size of the sample to minimize type I and type II error.* *Select " no" if minimum sample size and power level are not reported.* ***If no is selected, answer NA for B18*** *.*
B.5	*Select "yes" if the author stated that the subjects were randomly assigned to a group. If there is any restriction on randomization, it should be clearly described (eg, blocking, matching, or stratification). Acceptable examples of randomization that minimize bias include computer generated assignment, coin toss, shuffling cards, and throwing dice.* *Select "no" if the author used methods that lead to greater bias, such as use of odd/even date of birth, date of admission, or medical record number.*
B.6	*Select "yes" if the author described and demonstrated a process in which the individual who determines the participant's eligibility for inclusion in the study is unaware of the group to which the subject will be assigned.* *Select "no" if the author did not clearly describe how the subjects are assigned to a group or whether the 'assigner' is not blinded .*
B.7	*Select "yes" if a control group of 'no-treatment' or a placebo group is used. This may include the use of a crossover design, control phase, or wait list, as long as one group receives only pre-post measures for an initial phase without intervention. However, if the natural history of the condition has been studied and shown to be worse than a standard of care, a comparison group using the same standard of care can be substituted for the control group. The standard of care group should receive equivalent time and attention as the other treatment groups.* *Select "no" if a control group of 'no-treatment' or a placebo group is not used and the natural history has not been studied and reported.*
B.8	*Score "yes" if the subjects enrolled in the study are blinded from knowing whether they are in a group receiving a placebo, the standard of care, or an experimental treatment. (Note: This is separate from the allocation process.)* *Select "no" if the subjects are not blinded to their assigned intervention.*
B.9	*Select "yes" if the study references or describes the frequency, dosage, and other details sufficient for assessment, setting description, and reproducibility for all interventions used in the study.* *Select "no" if all study groups are not sufficiently described to reproduce the study.*
B.10	*Select "yes" if the study describes acceptable levels of attrition, unplanned crossover, subject compliance with protocol, and any other conditions that may contaminate the intervention allocation by the subject or researcher, such that your confidence in the results are not compromised.* *Select "no" if the attrition, crossover, or non-compliance with the protocol significantly compromises your confidence in the results or if breeches to the study protocol are not described.*
B.11	*Select "yes" if the author stated whether or not adverse events occurred.* *Select "no" if the author did not state whether or not adverse events occurred.*
B.12	*Select "yes" if the authors state that they receive no financial benefits or have personal relationships with sponsoring institutions or products used in the study, or if the disclosures of any conflicts do not compromise the study's research methods or results. Disclosures should include grant sources for the research.* *Select "no" if the authors reveal financial benefits or personal relationships with sponsoring institutions or products used in the study that can bias the study's results.* *Select "Not declared" if there is no mention of conflicts of interest by the authors. Do not award points for this selection.*

Last Updated: 1/30/16
Contact: practice-dept@apta.org

Instructions: B. Quality Evaluation-Outcome of Interest	
B.13	*Select "yes" if the validity has been established for this outcome measure in one of the of following ways: relevant prior studies are cited or validity was established within the study. Examples of dimensions of validity include face, content, concurrent, criterion, predictive, and construct validity.* *Select "no" if the outcome of interest has not been validated, or evidence provided is not sufficient to support validity for this study.*
B.14	*Select "yes" if reliability has been established for the outcome measure in one of the following ways: relevant prior studies are cited or reliability was established within the study. These may include test-retest, intra-tester, and/or inter-tester reliability as appropriate to the study. A range of recommended thresholds is between 0.7-0.9 but may be higher depending on the type of measure, the sample size and how results will be used.* *Select "no" if the reliability of the outcome measure is not describ ed, or they report reliability levels insufficient for the study.*
B.15	*Select "yes" if the reliability level is reported for measures that are performed by the study's tester(s). (Note: The intra- or inter-reliability level is typically described as a reliability coefficient.)* *Select "no" if the reliability level of the study's tester(s) is not reported.* *Select "N/A" if the outcome uses a measure that is not performed by a tester on a subject (eg, self report).*
B.16	*Select "yes" if the individual(s) who performed the testing on the subjects, and/or data collectors who harvest data from various sources, are blinded to the treatment group.* *Select "no" if the study does not indicate whether or not the testers or data collectors are blinded.*
B.17	*Select "yes" if intention-to-treat was reported. Intention-to-treat analysis means that subjects were analyzed within the groups that they were originally assigned to regardless of their adherence. The intention-to-treat analysis provides the conservative estimate of effect because it includes contamination from unplanned crossover and dropout. In the case of missing data, the researchers should specify the methods used for imputation of missing data. Regardless of how calculated, the original number of enrolled subjects must be included in the final data se* *Select "no" if only the as-treated analysis was reported. As-treated analysis means the subjects are analyzed based on treatment that they actually received. The as-treated analysis is likely biased toward an inflated estimate of effect.* *The 2 analyses provide the range of effect, with the true effect likely between the 2.* *Select "N/A" if there is no comparison group (e.g. single subject design).*
B.18	*Select "yes" if the sample size met the power requirement. (Note: the power requirement should be reported [see B .4]).* *Select "no" if the reported sample size did not meet power requirements .* *Select "N/A" if power analysis was not reported.*
B.19	*Select "yes" if the rater is confident that the groups' outcomes would not be affected by baseline differences between study groups.* *Select "no" if the rater is not confident that the groups' outcomes would not be affected by baseline differences between study groups.* *Select "N/A" if there is no comparison group (e.g. single subject design).*
B.20	*Check "yes" if sufficient data is reported to calculate the reported confidence interval, or an established MDC is referenced, and effect sizes are appropriately reported. [Note: Confidence intervals typically require number of groups, sample size, observed means, and standard deviations (eg, with continuous data); effect sizes are typically reported as a correlation, odds ratio, or similar calculation.]* *Select "no" if sufficient data to calculate the reported confidence interval and/or effect sizes are not reported.*

Quality Ratings - *use the following definitions to determine the overall quality of the study based on combined strengths/weaknesses of the design and the individual outcome.*	Studies are assigned 1 of 4 overall quality ratings based on the critical appraisal tallies for overall design and the individual outcome measure. • High quality studies are typically randomized clinical trials with greater than 80% follow-up, blinding, and appropriate randomization procedures. • Acceptable quality (the study does not meet requirements for high quality and weaknesses limit the confidence in the results). • Low quality: the study has significant limitations that substantially limit confidence in the results. • Unacceptable quality: serious limitations—exclude from consideration in a guideline or as sole support for patient management.

Tally Calculation:

Design Tally: One ('1") point is given for each yes and zero("0") points for each "no." Sum of scores for items B1-B12.

Outcome Tally One ("1") point is given for each "yes." Zero ("0") points for each "no" and "N/A." Sum of scores for items B1-B20.

Notes: C. Impact	
C.1	*Possible explanation: Interpretation is consistent with results, balancing benefits and harms, and considering other relevant evidence.*
C.5	*Does the study measure the outcome at a time appropriate for capturing the intervention's effect?*

Last Updated: 1/30/16
Contact: practice-dept@apta.org

APTA Critical Appraisal Tool for Experimental Intervention Studies Responses

Inclusion: Randomized controlled trial (Includes 'no-treatment' group) ☐ Quasi-expeimental (no randomization; with or without true control ☐
Randomized clinical trial (Includes a comparison group receiving placebo/sham or standard of care) ☐ Single subject design (must have a minimum of "a" and "b" phase) ☐
Other experimental study: _____

Exclusion: Non-experimental studies (eg, Cohort, Case series) ☐

Title: _____

Citation: *Authors:* _____ , *Journal:* _____, *Volume:* _____ , *Year:* _____

Reviewer: _____ **Date:** _____

A. Study Question and Design			
	Before evaluating and scoring the quality		**Reviewer Comments:**
	A.1	What is the primary purpose of the study?	***P** opulation:*
			***I** ntervention method(s):*
			Dosage Intensity / *Tx Frequency*
			***C** omparison:*
			Dosage Intensity / *Tx Frequency*
			***O** utcome(s):*
			***T** imeframe:*
	A.2	What is the study setting?	*Study setting. (Check all that apply)* ☐Laboratory ☐Acute Care ☐Inpatient Rehab ☐Outpatient Clinic ☐Home ☐Community (eg, community center) ☐Other _____
			Geographic setting. ☐ Urban ☐Suburban ☐Rural ☐Unknown
			Describe other factors relevant to the study questions (eg, country, health care delivery system):
	A.3	Did the study author indicate it was in compliance with IRB regulations?	☐ Yes ☐ No

Responses to the following items are used to determine quality rating(s): This critical appraisal tool evaluates the overall design separately from the outcomes of interest. The results of the overall design, added to the individual outcome of interest tally, are used to inform the quality rating for each outcome.

When this form is used for critical appraisal by multiple reviewers, a consensus should be reached by the reviewers as to the primary outcome(s).

B. Quality Evaluation

Overall Design (Select "yes" or "no" for each question about the overall design of the study.)

B.1	Are inclusion criteria for subjects specifically described?	☐ **Yes** ☐ **No**
B.2	Are exclusion criteria for subjects specifically described?	☐ **Yes** ☐ **No**
B.3	Do recruitment methods reasonably attempt to minimize bias in the sample?	☐ **Yes** ☐ **No**
B.4	Was an a priori power analysis used to determine the minimum sample size?	☐ **Yes** ☐ **No**
B.5	Were participants randomly assigned to intervention groups?	☐ **Yes** ☐ **No**
B.6	Was allocation to the treatment group concealed?	☐ **Yes** ☐ **No**
B.7	Was there a control group?	☐ **Yes** ☐ **No**
B.8	Were the subjects blinded to the treatment group?	☐ **Yes** ☐ **No**
B.9	Are the interventions for all study groups described in sufficient detail to be reproducible?	☐ **Yes** ☐ **No**
B.10	Was adherence to the study protocol described?	☐ **Yes** ☐ **No**
B.11	Was the presence or absence of adverse events described?	☐ **Yes** ☐ **No**
B.12	Are conflicts of interest managed to enhance confidence that the results are not biased?	☐ **Yes** ☐ **No** ☐ **Not declared**
	Design Tally (Yes=1 and No=0):	_____

Last Updated: 1/30/16
Contact: practice-dept@apta.org

B. Quality Evaluation

Outcomes of Interest (Select "yes," "no," or "N/A for each outcome of interest.)

Identify 5 outcome(s) of interest that address your purpose for reading the study. Use additional pages if more than 5 outcomes are of interest.

Outcome 1: _____ **Measure used:** _____

Outcome 2: _____ **Measure used:** _____

Outcome 3: _____ **Measure used:** _____

Outcome 4: _____ **Measure used:** _____

Outcome 5: _____ **Measure used:** _____

		Outcome 1:	Outcome 2:	Outcome 3:	Outcome 4:	Outcome 5:
B.13	For each outcome of interest, was a measure chosen that has established validity?	☐ Yes ☐ No	☐ Yes ☐ No	☐ Yes ☐ No	☐ Yes ☐ No	☐ Yes ☐ No
B.14	For each outcome of interest, was a measure chosen that has established reliability?	☐ Yes ☐ No	☐ Yes ☐ No	☐ Yes ☐ No	☐ Yes ☐ No	☐ Yes ☐ No
B.15	Was the reliability level of the tester(s) reported for each outcome of interest?	☐ Yes ☐ No ☐ N/A	☐ Yes ☐ No ☐ N/A	☐ Yes ☐ No ☐ N/A	☐ Yes ☐ No ☐ N/A	☐ Yes ☐ No ☐ N/A
B.16	Was the tester(s) and/or data collector(s) blinded to the treatment group(s) of interest?	☐ Yes ☐ No	☐ Yes ☐ No	☐ Yes ☐ No	☐ Yes ☐ No	☐ Yes ☐ No
B.17	Were subjects analyzed in the groups to which they were assigned, e.g., intention-to-treat analysis?	☐ Yes ☐ No ☐ N/A	☐ Yes ☐ No ☐ N/A	☐ Yes ☐ No ☐ N/A	☐ Yes ☐ No ☐ N/A	☐ Yes ☐ No ☐ N/A
B.18	Did the sample size meet power requirements for the outcome of interest?	☐ Yes ☐ No ☐ N/A	☐ Yes ☐ No ☐ N/A	☐ Yes ☐ No ☐ N/A	☐ Yes ☐ No ☐ N/A	☐ Yes ☐ No ☐ N/A
B.19	For each outcome of interest, are the study groups equivalent at baseline?	☐ Yes ☐ No ☐ N/A	☐ Yes ☐ No ☐ N/A	☐ Yes ☐ No ☐ N/A	☐ Yes ☐ No ☐ N/A	☐ Yes ☐ No ☐ N/A
B.20	For each outcome of interest, is the estimated effect size and its precision reported?	☐ Yes ☐ No	☐ Yes ☐ No	☐ Yes ☐ No	☐ Yes ☐ No	☐ Yes ☐ No
	Outcome tally:	Outcome #1 _____	Outcome #2 _____	Outcome #3 _____	Outcome #4 _____	Outcome #5 _____
	Design Tally: (questions B1-B12)			_____		

Determine the overall quality of the study based on the combined strengths and weaknesses of the Design and the individual Outcome.

	Outcome 1:	Outcome 2:	Outcome 3:	Outcome 4:	Outcome 5:
Quality Rating: (See instructions for suggested rating definitions.)	☐ High Quality Study	☐ High Quality Study	☐ High Quality Study	☐ High Quality Study	☐ High Quality Study
	☐ Acceptable Quality Study	☐ Acceptable Quality Study	☐ Acceptable Quality Study	☐ Acceptable Quality Study	☐ Acceptable Quality Study
	☐ Low Quality Study	☐ Low Quality Study	☐ Low Quality Study	☐ Low Quality Study	☐ Low Quality Study
	☐ Unacceptable Quality	☐ Unacceptable Quality	☐ Unacceptable Quality	☐ Unacceptable Quality	☐ Unacceptable Quality

C. Impact

	The following items are not scored but are useful to help determine the impact of the study:	**Reviewer Comments:**
C.1	Is the conclusion justified given the conduct of the study (eg, sampling procedure, measures of outcome used, and results achieved)?	
C.2	Were limitations of the study reported?	
C.3	Could interventions be applied in usual practice?	
C.4	Are there cost considerations to this intervention that might impact clinical utility?	
C.5	Given the purpose of the study, was follow-up sufficiently long?	

Comments: