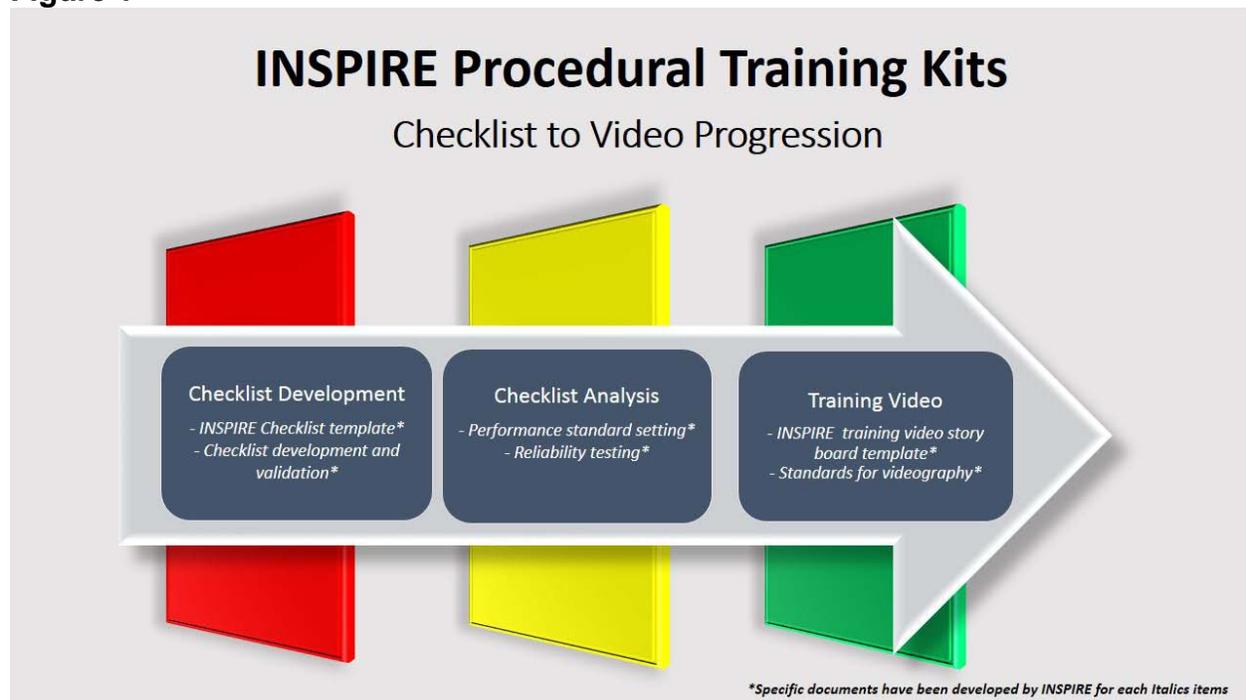




Why checklist development and validation first?

Development of a procedural checklist that has strong evidence of validity is the first step in the pathway to developing an INSPIRE procedural training 'kit', which includes a procedural checklist, a training video and train-the-trainer (TTT) materials. The process for developing a INSPIRE procedural training 'kit' (excluding the TTT materials) is outlined in Figure 1. Creating a checklist with strong evidence of validity first, ensures that the steps for performing the procedure described in the video (which correspond to the checklist) are valid, accurate and comprehensive. In this document we describe the process recommended by INSPIRE to develop and collect evidence of validity for a procedural skills checklist.

Figure 1



Checklist Development:

Step -1 – Review the INSPIRE Procedural Skills Checklist Template

INSPIRE has developed a general template for procedural checklists which should be used to develop an INSPIRE procedural checklist.

Step 2- Review of the medical literature for similar checklists and references

When beginning to develop a procedural checklist, a thorough search of medical literature should be undertaken in order to identify pre-existing published checklists or other relevant references on performing the procedure. If a validated checklist is provided in a publication it should be reviewed in detail and referenced. Any publication/presentation of your newly developed checklist should clearly describe how your checklist is different from the source checklist.



INSPIRE Procedural Checklists Development and Validation Instructions

Materials to Search:

Peer-reviewed Journals: A comprehensive literature review should be conducting using the Ovid MEDLINE and PubMed databases to identify English language articles published on checklists used for teaching or measuring competence in the procedure. A combination of relevant medical subject headings and/or keywords should be used in the search strategies. Search dates should span at least 20 years (1993 to 2013). This will result in an initial list of articles for review. Investigators should then review all titles for possible inclusion, and determine which warrant further review based on abstracts/papers. The reference sections of these articles should also be reviewed for identification of other relevant articles.

Textbooks/Procedure Guides: Published text books on procedures should be referenced in the development of the checklist. Such texts often include detailed lists of the steps of the procedure and relevant information on the indications, contraindications and anatomy.

Grey literature: Talking to colleagues and searching for online content using Google, or other search engines, can reveal previously developed checklists that were never formally published. INSPIRE also has a repository unpublished checklists that are available for reference.

MedEdPORTAL: The general MedEdPORTAL library and the Directory and Repository of Educational Assessment Measures (DREAM) collection (<https://www.mededportal.org/about/initiatives/dream/>) should also be searched for checklists on procedural training and assessment.

Step 3- Initial checklist draft creation

Based on the findings and available references in *Step 2* the authors should develop a *draft checklist* using the INSPIRE Procedural Checklist Template. It is critically import to only include on the checklist those steps/metrics that are easily observable and can be measured. All steps/metrics should be able to be defined as follows: 1) done independently and correctly without prompts, 2) done correctly but required prompts, 3) done incorrectly, or 4) not done at all. This step is vital to increase the reliability of the checklist.

Step 4- Checklist draft revision using the Delphi method

A Delphi method should be used to revise the draft checklist into a final format. The Delphi method involves an iterative process of review by a panel of experts that is designed to achieve consensus regarding the final list of checklist steps/metrics.

Step 4a- Identifying “Experts”

An ‘expert’ can be defined as a board certified physician in a specialty relevant to the procedure, who has performed the procedure themselves with success on multiple occasions, and who has served as a primary supervisor for the procedure. A total of **8 to 10 experts** should be sought for the Delphi process. In some procedures, it may be beneficial to include experts from multiple subspecialties (i.e. PEM, PICU, NICU, etc.) and multiple fields (anesthesia, surgery, etc.) to expand the breadth of expertise.

Step 4b-Delphi Rounds

During the Delphi rounds, experts will review the steps/metrics on the draft checklist. They should be instructed to rate the importance of performing each item on the checklist on a **7-point scale**, where 1, 2, and 3 represent not important, 4, 5, and 6 represent somewhat important, and 7 represents mandatory. Experts will be asked to score each item, revise items, write in comments, and suggest new items. The checklists with comments should then be returned to a working group member for data summarization. During the Delphi process the experts should be blinded to the input of other experts to prevent any single expert from unduly influencing the group’s decisions about items. Experts should also not discuss their impressions of the checklist items with the other experts. One working group member will remain blinded to the identities of the experts and will serve as a blinded reviewer of scores and edits during the Delphi process.



INSPIRE Procedural Checklists Development and Validation Instructions

The mean score for each steps/metrics should be calculated. Items on the checklist with a **mean rating** of 1 to 3 (*not important*) should be *removed* from the checklist. Revisions should be made based on the free-text edits and comments provided by the experts. The revised checklist should then be re-distributed to the experts for a second round of rating and editing. Again, the mean score for each the steps/metrics should be calculated for Delphi round 2. This process will continue until all experts agreed on the checklist items (e.g. **all items have an average scores ≥ 4**), and no additional revisions are needed, based on the experts' comments. At this point, the checklist contents will be finalized.

Summary of Steps in Delphi Rounds:

1. During this process, each "expert" will individually review the steps on the draft checklist.
2. Experts will rate the importance of performing each item on the checklist on a **7-point scale**, where 1, 2, and 3 represent *not important*; 4, 5, and 6 represent *somewhat important*; and 7 represents *mandatory*.
3. In addition to providing a numerical score for each checklist item, each expert should revise existing items, write in comments, and suggest new items as needed.
4. The checklists with comments should then be returned to lead investigator for data summarization.
5. During the Delphi process the experts will be blinded to the input of other experts to prevent any single expert from unduly influencing the group's decisions about items. Experts should also not discuss their impressions of the checklist items with the other experts. One working group member will remain blinded to the identities of the experts and will serve as a blinded reviewer of scores and edits during the Delphi process.
6. The mean score for each step will be reviewed, and items with low scores will be removed.
7. Revisions to the checklist will be made based upon free-text edits and comments provided by the experts.
8. The newly revised checklist will be re-distributed to the experts for a second round of rating and editing, followed by
9. This process will continue until all experts agree on the checklist items and no additional revisions are needed based
10. At this point, the checklist will be finalized and is ready for reliability testing.

Step 5- Statistical Analysis

The final experts' ratings on the checklist should be used to calculate internal consistency of the checklist using Cronbach's α . A Cronbach's α between 0.7 and 0.9 is desired. This statistical analysis can be easily performed using a free trial version of *MedCalc* software (http://www.medcalc.org/manual/cronbach_alpha.php) To evaluate agreement amongst the experts Cohen's Kappa can be determined, based on expert scores for each Delphi round.



Evidence of Checklist Validity:

What is 'validity?'

Checklist validity refers to whether or not the checklist's scores accurately reflect the construct it is intended to measure (e.g. the ability to perform the procedure). Validity is never absolute, but rather measured in degrees, based on the available evidence from various sources. Thus, instead of stating a checklist is 'valid', a more accurate description would be that the data derived from the checklist had 'strong evidence of validity', based on several lines of evidence, which are discussed below.

In general, evidence of validity is categorized into five sources, which include: *Content*, *Response Process*, *Internal Structure*, *Relationship to other variables*, and *Consequence*. A list of these along with example methods to provide evidence of validity for each is provided in Figure 2.

1. **Content validity** is the basis for all validity, and refers to whether or not the content (e.g. steps/metrics) in your checklist match the construct that the tool is intended to measure. For example, does completion of all the items on the procedural skills checklist accurately reflect the ability to competently perform the procedure? Do the steps on the procedural checklist include those needed for a competent practitioner to successfully complete the procedure? Are there extraneous steps? Are there missing steps? Evidence of content validity can be obtained by basing the steps/metrics included on checklist on previously published tools from the medical literature, and/or textbooks/procedure guides. The work done in developing the checklist using the Delphi method above contributes to content validity.

2. **Response process validity** refers to evidence supporting that the sources of error in administering the checklist have been minimized. For example, have trainers been instructed on the use of the checklist, and is there consistency in how assessment or evaluation is completed? To provide evidence of response process validity, measures should be taken to protect the integrity of the data provided by the checklist by controlling for potential sources of error associated with the use of the checklist, including rater training and standardizing the environment during checklist use, as much as possible. We hope to increase response process validity by develop train-the-trainer materials to standardize the administration of the checklist.

3. **Internal structure validity** relates to the statistical reliability of the checklist. The Cronbach's alpha determined during Delphi testing contributes evidence to internal structure validity. Other specific measures include the reproducibility of scores of the checklist (e.g. inter-rater reliability) and how well the checklist discriminate between high and low performers. For details on reliability testing please refer to INSPIRE Procedural Checklist Reliability Testing Instructions.

4. **Relationship to other variables validity** refers to the concept that one's overall score on the procedural checklist should correlate with other established methods of measuring procedural success, or competency. Such evidence can be provided based on evidence that higher scores on the checklist correspond to advanced training levels, or correlation of scores with other performance assessments (e.g. global skills ratings, 360° evaluations, OSCE, etc.).

5. **Consequence validity** involves the impact of the checklist's scores on trainees, teachers, patients, and society. An important aspect of consequence validity is the type of assessment the checklist is used for; formative or summative. If summative, what are the consequences of 'failing' on the checklist? Will a trainee be prohibited from performing a procedure on a patient without 'passing' the checklist on a simulator? If the checklist is used for summative assessment, then a defensible method must be used to determine the lowest passing score. For details on reliability testing please refer to INSPIRE Procedural Checklist Performance Standard Setting Instructions. Another important aspect of consequence validity is that the procedural checklist score should be able to determine a meaningful clinical consequences, e.g. pass or fail cutoff scores on the checklist should differentiate between different clinical outcomes (e.g. procedural success and failure). Evidence for this category of validity will also likely be developed when we employ the checklist in both the simulated and clinical settings.



Figure 2. Evidence of Validity for Procedural Skills Checklists	
Types of validity evidence	Methods to provide evidence of validity
Content Validity	<ul style="list-style-type: none"> ◆ Conduct literature review and develop checklist based on other published checklists, textbook descriptions, etc. ◆ Evaluation of checklist by subject matter experts using Delphi method
Response Process	<ul style="list-style-type: none"> ◆ Perform proper rater training ◆ Standardizing the simulation environment by using the same simulator, equipment and set up
Internal Structure	<ul style="list-style-type: none"> ◆ Compare checklist scores of experience to less-experienced providers (e.g. intern vs. resident vs. fellow vs. attending) ◆ Perform Cronbach's alpha
Relation to Other Variables	<ul style="list-style-type: none"> ◆ Correlate checklist performance to other tests of competency or expertise (e.g. PGY status, frequency of procedures, years of service, patient care ratings on evaluations, etc.) ◆ Correlation of checklist scores to global rating scale scores or other methods of performance assessment.
Consequence Validity	<ul style="list-style-type: none"> ◆ Is the checklist used for formative or summative assessment? ◆ What are consequences of failing to obtain a passing score on the checklist? ◆ Correlate checklist scores to simulated or clinical success with procedure ◆ Evaluate <i>predictive validity</i> of checklist scores on a simulator with clinical success at the bedside

References:

1. Downing SM. Validity: on meaningful interpretation of assessment data. *Med Educ* 2003;37:830
2. Cook D, Beckman T. Current concepts in validity and reliability for psychometric instruments: Theory and application. *Am J of Med.* 2006;119:e7-e16.
3. Messick S. *Validity*. In: Linn RI, ed. *Educational Measurement*, 3rd ed. New York: American Council on Education and Macmillan; 1989.
4. MedEdPORTAL Assessment Instrument Guidelines. Updated October 3, 2011. Accessed 10/11/2013. Available at: <https://www.mededportal.org/download/261990/data/assessmentguidelinestemplate.pdf>
5. Feldt LS, Brennan RL. *Reliability*. In: Linn RL, editor. *Educational Measurement*, 3rd Ed. New York: American Council on Education and Macmillan; 1989.
6. Berg D, Berg K, Riesenber LA, Weber D, King D, Mealey K, Justice EM, Geffe K, Tinkoff G. The Development of a Validated Checklist for Thoracentesis : Preliminary Results *Am J Med Qual.* 2013;28(3):220-6



INSPIRE Procedural Checklists Development and Validation Instructions

7. Berg K, Riesenber LA, Berg D, Schaeffer A, Davis J, Justice EM, Tinkoff G, Jasper E. The Development of a Validated Checklist for Radial Arterial Line Placement: Preliminary Results. *Am J Med Qual*. 2013 Jul 11. [Epub ahead of print]
8. Riesenber LA, Berg K, Berg D, Davis J, Schaeffer A, Justice EM, Tinkoff G. The Development of a Validated Checklist for Femoral Venous Catheterization: Preliminary Results. *Am J Med Qual*. 2013 Sep 17. [Epub ahead of print]
9. Riesenber LA, Berg K, Berg D, Mealey K, Weber D, King D, Justice EM, Geffe K, Tinkoff G. The development of a validated checklist for paracentesis: preliminary results. *Am J Med Qual*. 2013;28(3):227-31.
10. Berg K, Riesenber LA, Berg D, Mealey K, Weber D, King D, Justice EM, Geffe K, Tinkoff G. The development of a validated checklist for adult lumbar puncture: preliminary results. *Am J Med Qual*. 2013;28(4):330-4.
11. Riesenber LA, Berg K, Berg D, Schaeffer A, Mealey K, Davis J, Weber D, King D, Justice EM, Geffe K, Tinkoff G. The development of a validated checklist for nasogastric tube insertion: preliminary results. *Am J Med Qual*. 2013;28(5):429-33.
12. Huang G, Newman L, Schwartzstein R, et al. Procedural Competence in Internal Medicine Residents: Validity of a Central Venous Catheter Insertion Assessment Instrument. *Acad Med*. 2009; 84:1127–1134.
13. Tavakol M, Dennick R. Making sense of Cronbach alpha. *Int J Med Educ* 2011;2:53-55