ASC
FOR GOOD MEASURE

DEVELOPING AN ACCREDITED CERTIFICATION

WHITE PAPER

Developing an accredited certification program is no easy feat, requiring at least a year of effort from a team of experts and tens (often hundreds) of thousands of dollars. You cannot simply get an existing test accredited. The testing program, as well as the entire organization, must be built from the ground up according to accreditation guidelines. However, the benefits are very real, as accreditation serves as a stamp of approval, facilitating the recognition of the certification as a benchmark in your field.

There are a wide range of standards that must be adhered to in order to achieve accreditation many of which have nothing to do with the test itself. These can include topics like board governance, organizational finance, education/training, and recertification. Here, we will focus on the psychometric aspects, as those are typically considered the most "black box."

Formal development of an accreditation-worthy certification test is not a linear process, but rather a cycle that requires planning ahead more than one revolution. The validity of a test score is supported, in large part, by the strength of the connections between the different steps of the cycle. A broad example of this cycle is below, with more detailed descriptions following.
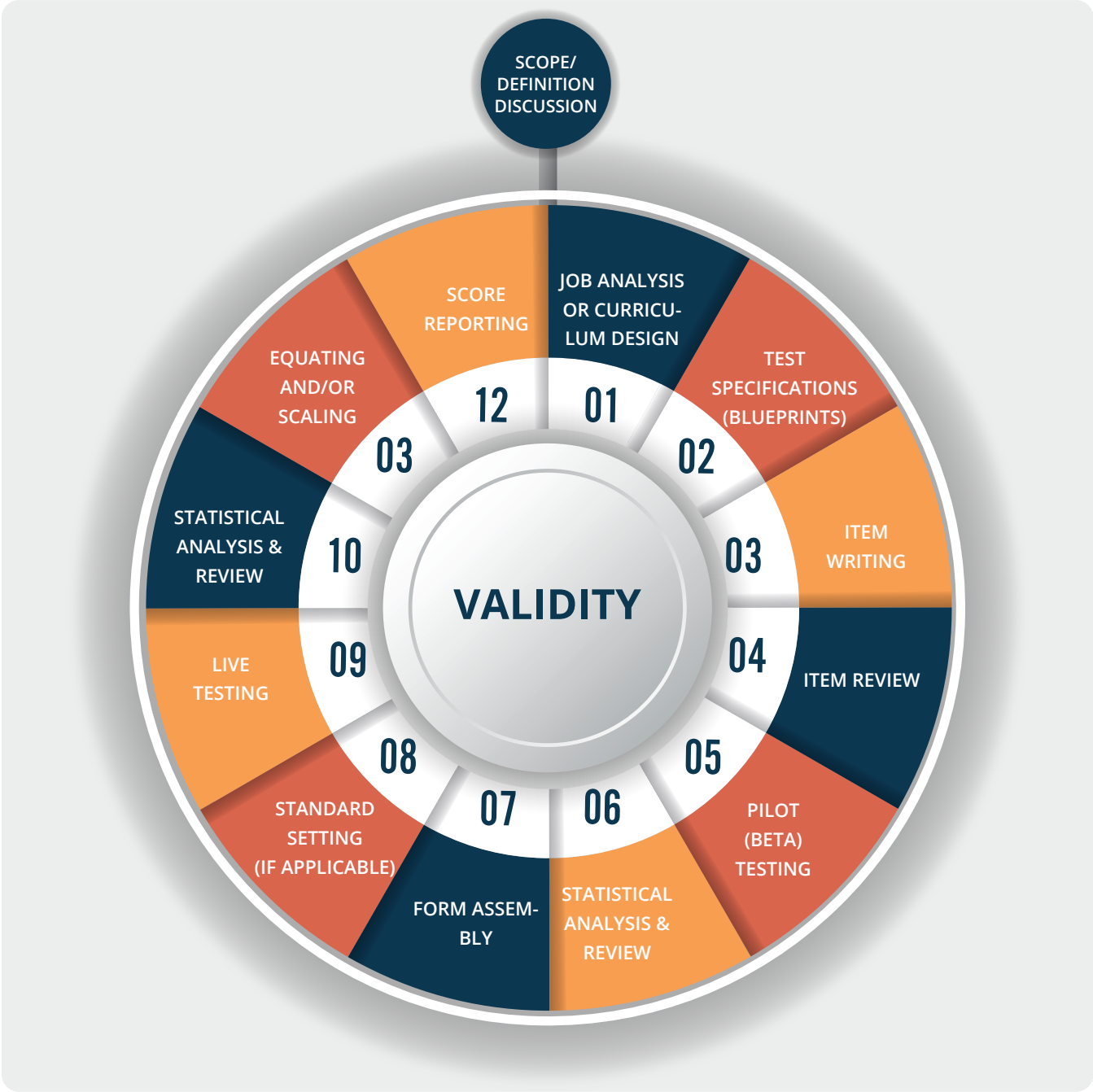
DETAILED
DESCRIPTION

Test Definition/Scope:
The first step is to define the goal of the test, its role in your profession, and what you want the credential to represent. This lays the foundation for validity, which is an accumulation of evidence that says the test scores mean what we want them to mean. If we don't say what we want them to mean, the rest of the process is unfounded.

Job Analysis:
The purpose of this step is to provide quantitative information regarding what knowledge, skills, or abilities (KSAs) are required to perform the job successfully. It then reasons that the test should test these KSAs if the purpose is to screen candidates that will not be successful. This step typically utilizes a task inventory survey: a committee of experts developed a list of professional tasks, and your field is surveyed on which are most important and frequent. This obviously involves substantial time and cost. The end result is data that will justify the content to be covered in the test.

Test Specifications (Blueprints):
This step converts the results of the previous step into an outline of the test (e.g., there will be 10 items in Content Domain 1...). It will involve statistical analysis by a psychometrician and discussions with the exam sponsors.

Item Writing:
The time and cost involved in this step are variable dependent upon several conditions, including the number and item writing skill of the item writers and the sophistication of the content. Obviously, items can be written much more quickly for elementary mathematics than for advanced medical or legal topics. Item writers should be trained in best practices if they do not have previous experience. The actual training is usually several hours, though it is typically combined with the actual item writing in one workshop that can last from one day up to several days. This step is much easier if a reference list has been identified beforehand.

Item Review:
Before expending the effort to pretest items with examinees, the items should be reviewed by additional item writers or experts, both for format and content. Content expertise is more valuable in this step; psychometric expertise is less important.

Beta Testing or Pretesting:
Items should be tried out on a sample of examinees to obtain statistics that allow the items to be examined more closely. For example, it might be discovered that examinees tend to all select the same incorrect answer. Psychometricians are not necessarily involved in the actual administration to pretest examinees, but have an extensive role in the next step.

Review Pretest Statistics:
The results of the pretest sample should be analyzed by a psychometrician, who will review the results for psychometric aspects such as item difficulty, item discrimination, and score reliability. Items with potential issues will be flagged, and an explanation provided if possible. Items are then jointly reviewed in another workshop by subject matter experts and the psychometrician, and items are retired, replaced, or revised as needed. Depending on the number and quality of the items, this can take several days' worth of work.

Standard-Setting:
The cutscore (pass/fail score) for a certification exam cannot be set arbitrarily at a round number like 70%. Instead, it must be criterion referenced, and set by consensus of subject matter experts, necessitating another workshop/meeting. Several methods exist, the most common being the "modified Angoff" method.

Form Assembly:
Green-lighted items are selected to be in the form(s) to be used for live administration. Statistics, content domains, and overlap should all be taken into account. For the first form of a new testing program, this is fairly simple. It is much more complex if there are to be four forms for an established testing program, each with a certain amount of overlap to last year's form, with a specified level of difficulty. In some cases, the forms are pre-equated.

Live Testing:

Test is administered to actual examinees, either in a time window or continuously.

Equating and Scaling:

If it is necessary to ensure that this year's scores are comparable to last years, a statistical adjustment process called equating must be completed. The general idea is that if this year's exam scores are lower but we know the examinees are just as able, scores should probably be adjusted upward because this year's test was more difficult. In certain cases, this can be done before the test is released, called "pre-equating." This step involves a technical analysis by a psychometrician.
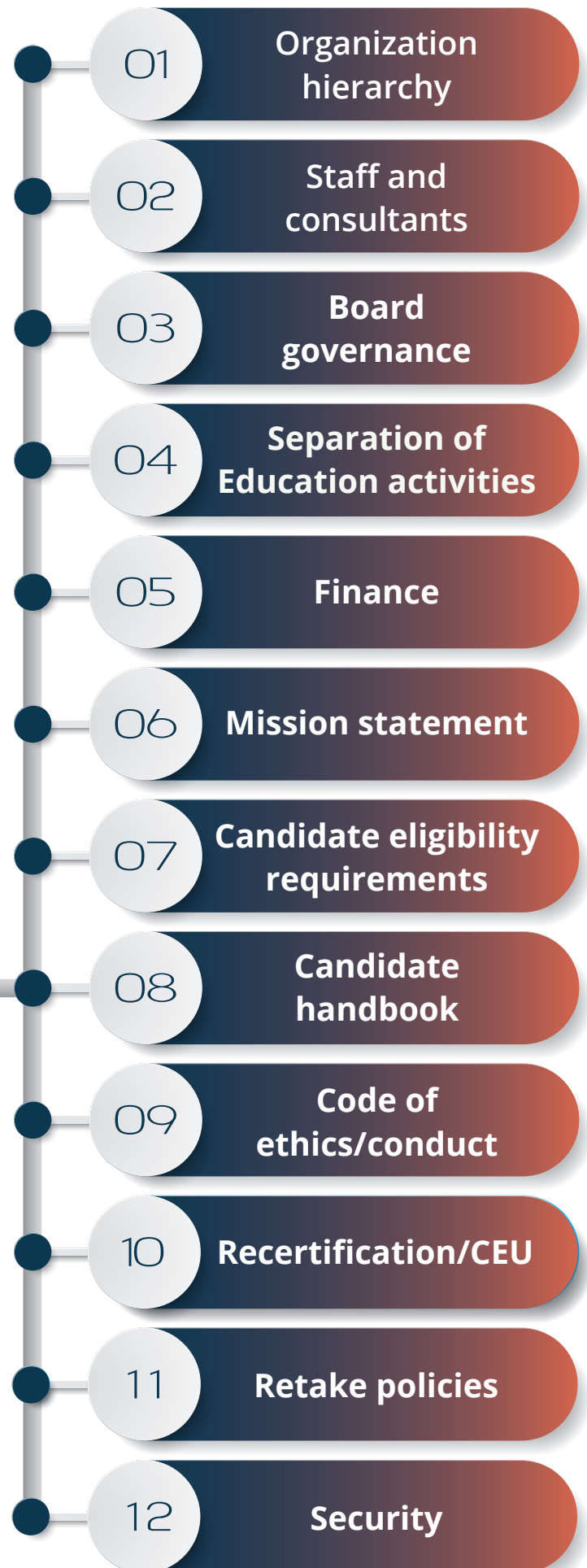
Score Reporting:

Scores are reported to examinees.

Annual statistical analysis:

Accreditation guidelines require a detailed technical analysis of exam results at least once per year. Again, this typically requires an analysis by a psychometrician and a workshop for experts to review the items and perhaps revise them for new test forms. However, ASC's innovative platform allows organizations to automatically produce their own reports.

THE BIG PICTURE

The process above is only the tip of the iceberg; the majority of accreditation standards are actually unrelated to the test itself, governing other aspects of the organization. If working towards accreditation, organizational staff can work on topics such as these while test development and psychometric personnel work on the tests.

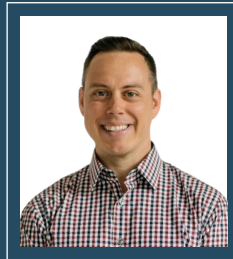|	
01	Organization hierarchy
02	Staff and consultants
03	Board governance
04	Separation of Education activities
05	Finance
06	Mission statement
07	Candidate eligibility requirements
08	Candidate handbook
09	Code of ethics/conduct
10	Recertification/CEU
11	Retake policies
12	Security

assess.com

IS IT **WORTHWHILE?**

Because the process of achieving accreditation is so lengthy and expensive, it is not always feasible. Before starting a new certification with the eventual goal of accreditation, it is essential to perform due diligence such as market research and SWOT analysis. Results of those allow for the evaluation of a compelling business case.

Nevertheless, the process of accreditation is often worthwhile, or it would not exist. Accreditation allows stakeholders in your industry to know that your organization upholds the highest standards and that your test has been built according to best practices. These, in turn, mean that the test is very accurate in identifying candidates with a recognized level of competence, the knowledge of which benefits both employers and consumers. In some cases, there are external reasons, such as the reduction in governmental funding opportunities of the program is not accredited.

ABOUT **ASC**

ASC (assess.com) provides online assessment software that is designed to help manage the entire test development and delivery process. Our platform leverages AI and automation to make it easier to develop high-quality exams and deliver with modern psychometrics like IRT and CAT.

NATHAN THOMPSON, PHD
CEO
ASSESSMENT SYSTEMS CORPORATION
solutions@assess.com

No review:
CATs rarely allow for examinees to return to items already administered, as the CAT has since adapted and it cannot unadapt.

Item exposure:
CATs are designed to select the best items in the bank, and these items will often become overexposed if a control algorithm is not implemented.

Recovery of poor starts:
CATs are susceptible to issues with examinee test anxiety, as the elimination of item review prevents someone from going back to the first few items. If they answered all those items incorrectly due to severe test anxiety, the test cannot correct itself.

Disadvantages of

CAT

Requirements:
CAT requires large sample sizes and extensive expertise. Testing programs need at least several hundred examinees, and an experienced Ph.D. psychometrician. Another white paper discusses the requirements to develop a CAT examination program, but some information is presented next.

Public relations:
Because of the complexity and the departure from the familiarity of the traditional exam paradigm, an organization must put forth more effort into public relations, explaining CAT and the reasons for using it.

Requires calibration:
Nonstandard item formats are often not easily scored in real time, or are not able to be calibrated with IRT. Both are required for CAT.