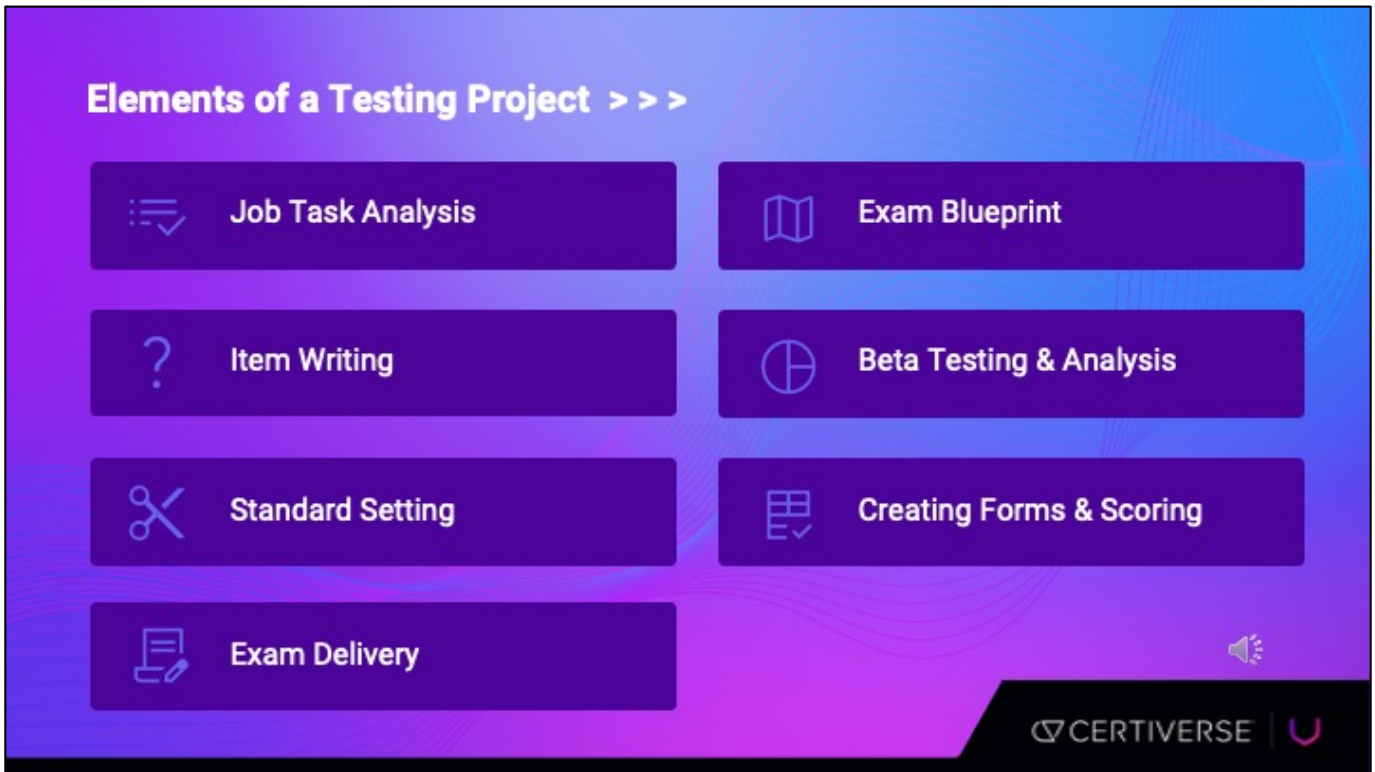


CU-101.2

Introduction to Testing:

# Elements of a Testing Project





In this video we'll be covering seven important elements of a testing project. The goal here is to simply introduce you to these elements and the purpose they serve. In later videos we'll dive deeper.


Let's get into the elements ...

- Sometimes called **Role Analysis** , **Job Analysis**, or similar, a **Job Task Analysis** – or **JTA** – is the process of identifying the Tasks that are performed by a person in a specific job or role.
- **Exam Blueprints**, or just Blueprints, at a minimum, describe the content areas to be covered in the exam and the relative importance or weight of those content areas. It

describes what the exam is intended to assess and measure.

- As I mentioned in a previous video, item is just a fancy term for a test question. **Item writing** is the process of authoring, reviewing, and editing that leads to the production of usable items.
- **Beta Testing & Analysis** is the process of administering items to real test takers for the purpose of collecting data about the performance of those items ... are they doing what we want them to do ... measuring what needs to be measured?
- **Standard Setting** is the process by which we determine the passing standard or cut score for an exam. Specifically, this the score above which test takers are judged to have "passed" the exam.
- A "Form" at its most basic is a specific set of items making up a version of the exam ... and often there is more than one Form to allow for enhanced security. In **Creating Forms & Scoring** we'll discuss how items are selected for placement on a form and also how we can use those items to calculate scores and generate outcomes.
- Finally, in **Exam Delivery** we'll talk briefly about some of the decisions that need to be made when it comes to how exams are administered to test takers.

**Documentation >>>**

 All elements of test development should be **Documented**

- > For **Test Takers** ... to ensure consistency, transparency, & fairness
- > For **End Consumers** [ Employers, General Public, etc.] ... to provide transparency & engender trust
- > For **Test Sponsors** ... to follow best practices, to ensure auditability and legal defensibility

>>> **Integrated** documentation is best! <<<

CERTIVERSE | U

Before we go into the specific elements, let's take a brief aside and talk about **Documentation**.

As we begin and progress through the elements of a testing project, it is very important to document both your overall approach as well as specifically what happens throughout the process.

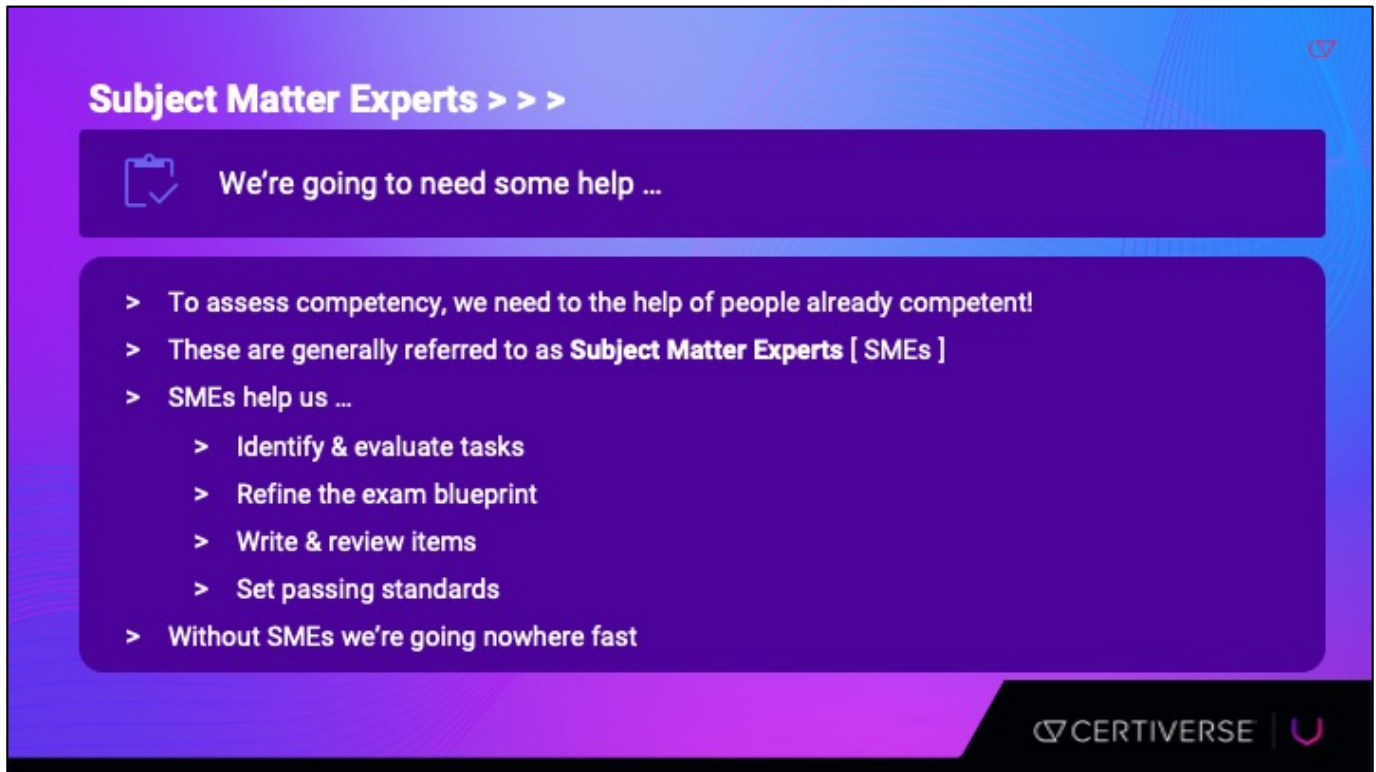
Why you ask?

- For **Test Takers** it ensures that there is clarity in what is expected of them in preparation and also what they can expect when being assessed.
- For **End Consumers** it provides transparency as to what any resulting credential means. For example, hiring employers want to know specifically what knowledge, skills, and


abilities are assessed. The general public wants to know that the folks they rely on to fly planes, perform surgery, and provide financial guidance - to name a few examples - are qualified to do so.

- Finally, **Test Sponsors** should want to assure that they are following best practices and are able to prove it if necessary. In some cases, they may even need to prove it in a court of law.


As a final comment before moving on ... my experience is that the best documentation results from using systems that "bake in" documentation or inherently document by enforcing process in software. Adopting a system that does this for you will increase efficiency dramatically.



**Subject Matter Experts >>>**

 We're going to need some help ...

- > To assess competency, we need to the help of people already competent!
- > These are generally referred to as **Subject Matter Experts [ SMEs ]**
- > SMEs help us ...
  - > Identify & evaluate tasks
  - > Refine the exam blueprint
  - > Write & review items
  - > Set passing standards
- > Without SMEs we're going nowhere fast

CERTIVERSE | 

Another quick but very important aside. A lot of work and expertise goes into developing a fair, valid, and reliable exam.

When we talk about assessing people for competence in a job, we need the expertise of the folks who do or know that job.

We call these people Subject Matter Experts, or SMEs for short.

SMEs will help us in many ways including ...

- Identifying the Tasks that make up a job
- Refining the content that needs to be assessed
- Creating the test questions

- ... and determining the level needed to pass each form

In short, SMEs are critical to making an exam happen

## Job Task Analysis >>>



An Exam should be based on a thorough analysis of curriculum, job, or role

- > It is critical to know what you're asking the exam to assess!
- > This the purpose of a **Job Task Analysis [ JTA ]**
- > A JTA is made up of three elements ...
  - > **Task Elicitation** – the **Tasks** performed by competent practitioners
  - > **Survey** – a broad review provides data to support task and their validity
  - > **Analysis** - judicious filtering and calculations identify critical tasks

So, you need to assess individuals for a job or a role ... where do you start? Well, we start by determining the demands of the job or role OR in the case of something like a general knowledge training course, what makes up the curriculum.

- This is critical! An exam that is not aligned to the job or role is not useful and can even be dangerous. We want to be sure a licensed pilot really knows how to fly that plane! To align the exam, we might think about this in different ways depending on the situation.

In the case of a job or role, we need to identify the critical **Tasks** performed by people competent in that job or role.

To operate a piece of equipment we might frame this in terms of the **Knowledge, Skills, and Abilities** – shortened to **KSAs** – needed to successfully, safely, and legally operate the equipment.



In the case of a college course, we might think about this in terms of **Learning Outcomes**, that when demonstrated represent mastery of the material being taught

We gather this information and validate it through a **Job Task Analysis**. To be clear, there a number of names for this element ... Job Analysis, Role Analysis, Task Analysis, etc.

Regardless of the name, the fundamental goal remains – to determine what is to be measured. I'll be consistently using the acronym JTA for the rest of this course. Similarly, going forward I'll be using the term **Task** consistently to refer to Tasks, KSAs, and/or Learning Outcomes.

In general, there are three main steps to executing a JTA ...

- **Task Elicitation** where we employ SMEs to gather as many of the tasks that make up the job as we can.
  
- Next we
  
  
  
  
  
  
  
  
  
  
- Finally we

## [ Exam ] Blueprint >>>



We need to know [ EXPLICITLY ] what the Exam is intended to assess

- > **Critical Tasks** are linked to **Domains** ... broad areas of knowledge/skill/content to be assessed
- > **Domains** and their relative weightings form the initial basis of the **Exam Blueprint**
- > [ Frequently ] domains are split or subdivided for more granularity/specificity
- > Relative importance of subdomains is established
- > **Domains** combined with weighting describes the distribution of items [ questions ] needed

>>> Methods for arriving at a final Blueprint can be highly variable! <<<

The purpose of an exam blueprint is to explicitly describe what is to be assessed and this must be linked back to the JTA in order to be defensible.

- As we just discussed, a successful JTA will have surfaced the critical tasks and provide insight into their relative weightings. Each task is generally linked to one or more content areas. which we refer to as **Domains**.

When we talk about Domains, we are describing broad content areas. For example, domains for an engineering role might include "Regulations", "Design Software", "Planning", and "Surveying".

When I mention "Regulations" in the context of engineering, most people will immediately have at least some abstract idea what we're talking about. On the other hand, SMEs should immediately know in detail what the domain

"Regulations" encompasses for their industry.

- Back to the JTA, from our analysis, we know the critical tasks and their weightings. This means we can also derive the critical Domains and their weightings. This provides us with an initial stab at the primary content areas to assess on the Exam.
- With those initial Domains and weightings in hand, the blueprint needs to be refined, with domains often being reorganized, subdivided, split or all of the above. This provides logical organization as well as more detail and granularity to clarify what is being assessed.
- These new domains and subdomains are themselves often weighted to identify their relative importance in the blueprint.
- The resulting set of domains & subdomains with their weighting describes the distribution of content needed to satisfy the blueprint ... and this will be instrumental in driving **item writing**

A final comment ... getting to a completed blueprint can be a highly variable process depending on the content being assessed. The team developing the assessment will need to react to and factor in the real-world constraints they encounter.

## Item Writing >>>



Driven by an exam blueprint, **Items** are written

- > Enough **Items** must be written [ **Authored** ] to satisfy the exam blueprint
- > Factors [ typically ] include ...
  - > Domain weighting ... determines distribution of items [ % ] by content area
  - > Length of exam [ in items ]
  - > Number of equivalent forms needed
  - > Allowances for "bad" items
- > Items should be authored following best practices and job norms
- > Items are [ typically ] **Reviewed** and **Revised** until they are **Accepted**

Let's talk about Item Writing. Again, the term **Item** just means test question. There are many different item types, especially since the advent of computer-based testing. However, the good old 4-option, multiple choice, single response format many of us have seen over our lives continues to be common and even dominant. So, for now, feel free to envision the 4-option multiple choice variant when you hear the term item.

- With our Blueprint in hand, we now know *the distribution* of items we need and the domains they will measure. We need to start writing – or authoring – those items! But how many items do we actually need?
- Here are some factors we need to consider ...
  - We need enough to satisfy the blueprint distribution by content area. But generally, this just tells us a percentage, not an absolute number.

- So, we need to factor in the length of the exam, often using the number of items. Will the exam be 50 items long? 100 items? Etc.
- We also need to factor how many equivalent forms or versions of the exam are we targeting. Do we need 2 versions? 3? This becomes a multiplier. Also, are we allowing or requiring overlapping items between forms? If so, we need to adjust for this factor.
- Finally, not all the items we author will end up performing well. Poorly performing items should not be used! So, we need to ensure we write enough extra items to account for this. In item writing, we think of this as a margin - or allowance - for bad items.
- Speaking of bad items, when we're writing items, we want to follow a process that maximizes the chance for as many good items, and as few bad items, as possible. This means we need to support our item authors in producing good items.

This includes exposing them to general best practices AND specific instructions and rules for item authoring relevant to the industry, job, and domain.

- Items are not just authored. They should also be reviewed by other SMEs, providing approval or feedback to the original author AND allowing for revisions as needed. Ultimately, multiple SMEs contribute to the production of each item.

Under a very simple scenario, one SME writes the item while one or more other SMEs review the item before approval.

The image is a screenshot of a presentation slide with a purple and blue background. At the top left, the title 'Beta Testing & Analysis' is followed by three right-pointing chevrons. Below the title is a dark purple rounded rectangle containing a white plus sign icon and the text 'Our new exam is **Delivered** and **Results** are **Psychometrically** analyzed'. Below this is another dark purple rounded rectangle containing a bulleted list of four items. At the bottom right of the slide, there is a speaker icon and the logo for 'CERTIVERSE'.

## Beta Testing & Analysis >>>

⊕ Our new exam is **Delivered** and **Results** are **Psychometrically** analyzed

- > We test the test!
- > This is called a **Beta [ Test or Exam ]**
- > Data is gathered and evaluated to assure exam is fit for purpose
- > To analyze data, we apply the appropriate **Psychometric** statistical techniques

🔊 CERTIVERSE

We've written a bunch of items that satisfy the blueprint AND that we think are good. However, we don't really know for a fact that these items are good.


In order to ensure a fair, valid and reliable exam empirically, we administer a Beta Test.

- Essentially, we "Test the Test"!
- We call this a Beta Test or Exam. Our Beta is made up of a set of items that satisfy the blueprint. Often this is most - or all - of the items we developed in the Item Writing phase.
- We administer these items to a set of test takers that are representative of the target population. We need a statistically significant number of test takers to ensure we get reliable data. Note that this number can vary dramatically depending on many factors.

- Once we have enough administrations, we perform psychometric analyses ... statistical analyses that provide insight into how items are performing. A couple of common examples include item "difficulty" – what percentage of test takers get the item correct. Another is item "discrimination" – a number that shows how well the item differentiates between low and high performing test takers. There are many other types of statistics that can be calculated at both the item- and form- level.

With this data in hand, we are ready to move on to the next step,  
Standard Setting

**Standard Setting >>>**

 **Standard Setting** determines the performance required to **Pass** the exam

- > We need to determine the **Passing Standard**, often referred to as the **Cut Score**
  - >>> **This determination cannot be arbitrary or capricious!** <<<
- > In standard setting, a single form is reviewed resulting in a passing standard for that form
- > There are a number of methodologies that can be used to arrive at the passing standard
- > **The Minimally Qualified Candidate**

CERTIVERSE | U

Standard Setting is the process used to determine the score required to “Pass” the test.

- This is often referred to as the **Passing Standard** or **Cut Score**.
  - The passing standard must be arrived at judiciously and by a consensus of SMEs ... contrary to what some of us learned in high school, the cut score is not magically always at 70%!
- To set a standard, we need a single form of the exam. The initial form generally “emerges” from the beta. For right now, we’ll gloss over the details of how that happens, but in a nutshell, the form should be composed of a set of items that performed well on the beta – we know this from our analyses - AND that together fully satisfy the blueprint.



- There are several recognized and accepted methodologies that are often used to arrive at a passing standard. The key is to select one and apply it appropriately.
- One concept that all standard setting methodologies rely on is called the **Minimally Qualified Candidate**.

Keeping in mind that our exam is attempting to establish competency, we need to make a binary decision ... competent or not competent.

To do this, we need to clearly understand what the metric is for competency. That is where the concept of the **Minimally Qualified Candidate** comes in. This is not a real person. It is a representation of the theoretical person who's just qualified - barely qualified – to do the job. Not over-qualified. Not highly qualified. Minimally qualified to enter the job or role.

This concept underpins any approach to standard setting. Your SMEs must understand and agree on who the minimally qualified candidate is and apply this understanding to setting the passing standard.

## Creating Forms & Scoring >>>



### Assembling forms and determining outcomes

- > A **Form** is a specific set of items that adheres to the blueprint
- > How forms are assembled can be highly variable
  - > Multiple forms are often used to increase security
  - > We want forms to be as similar as possible
- > **Scores** are the calculated outcomes of the test
  - > Outcomes may or may not be reported

>>> **Reliable outcomes are dependent on the design of the exam!** <<<





## Exam Delivery >>>



There are many considerations when it comes to **Exam Delivery**

How do candidates get access?

- > **Prerequisites/Eligibility & Booking** considerations
- > **Availability** by date, window, region, country, etc.
- > **Costs** to test
- > **Retake** policies

Under what conditions do candidates take tests?

- > **Paper/Pencil vs. Computer**
- > **Remote vs. In-Person**
- > **Proctored vs. Un-proctored**

Finally, let's talk very briefly about administering your exam. For Exam Delivery you'll need to make many decisions.

These decisions are necessarily informed by costs, technical constraints, exam design, security concerns, needed geographic reach, accessibility, and other concerns. With this in mind, here are a few of the things you'll need to consider ...

- You'll need to decide how, where, and when Test takers are able to access your exam.
  - Do test takers need to be authorized to test? For example, are there prerequisites or can anyone take the exam?  
Also, how will test takers book their testing session?
  - When and where will your exams be available? Will there be physical locations or are they accessed online?

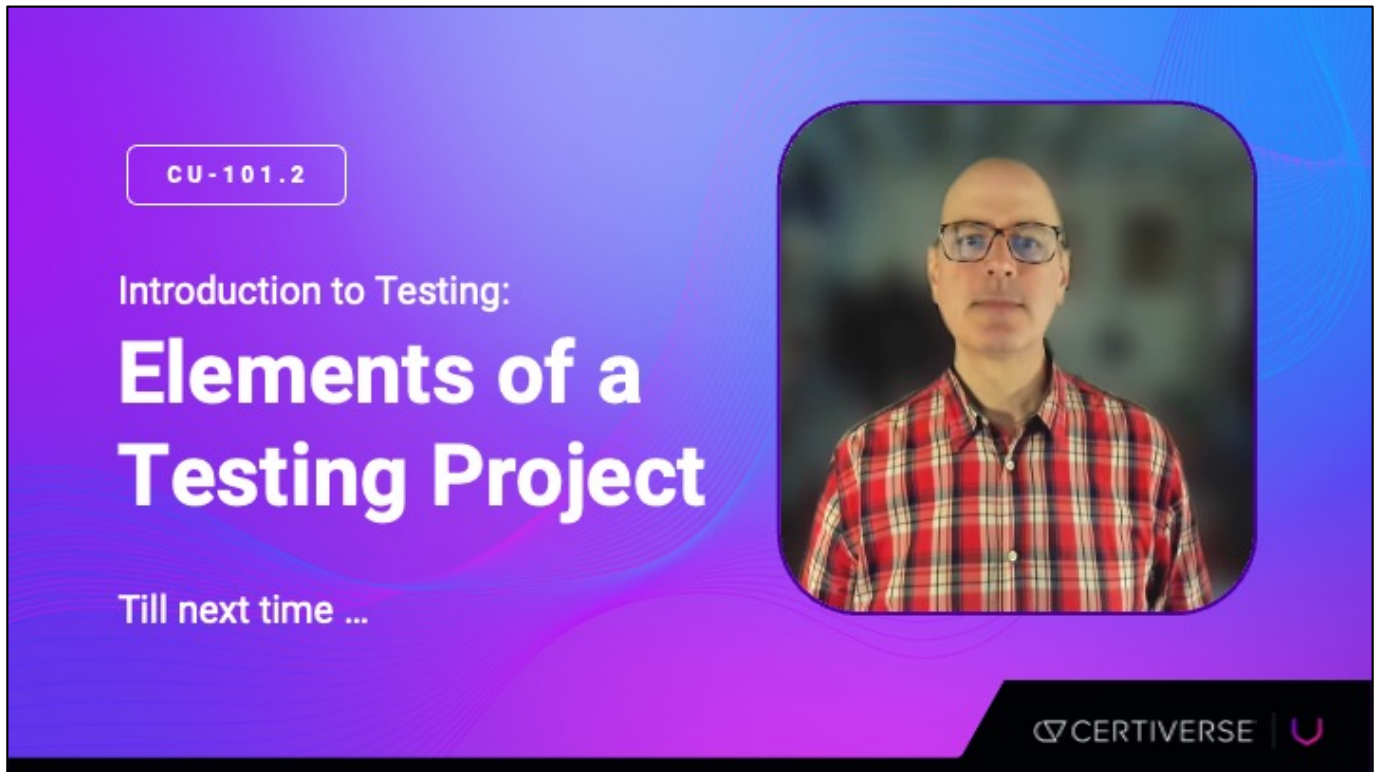
Will there be testing events or windows at specific times or is the exam available on-demand? Is access constrained by region or country?

- Are your tests free or do you collect payment? Are they bundled with training? Do you provide discounts to partners or for bulk purchases?
- Are candidates allowed to retake the test? If so, how often and in what windows?
- You also need to decide under what conditions test takers sit the exam ...
  - Are your assessments **Paper/Pencil** **Computer-based**?
  - Are they administered **Remotely** **In-Person**?
  - Are they **Proctored** **Un-proctored**?

These considerations are sometimes discrete and sometimes not! For example, you can deliver your exams remotely AND in-person.

On the other hand, in some cases, making one decision may preclude another. For example, for fairness, it's probably not a great idea to proctor some test takers and not others.

Ultimately, these decisions will need to be made holistically, factoring any tradeoffs.



CU-101.2

Introduction to Testing:  
**Elements of a  
Testing Project**

Till next time ...

CERTIVERSE | U

You can see that there are a lot of elements involved in developing and deploying a fair, valid, and reliable exam! Hopefully, you can also see that there's a well-worn path to get there.

In the next videos we'll be diving deeper into each of the elements discussed here so you get an even better understanding of what they involve.

That's it for this video! Thanks for watching ... we'll see you in the next!