psi Testing Excellence

Tell me more about...

Effective testing with psychometric principles

psiexams.com

01

Introduction

This guide pulls together insights from the application of psychometric principles to successful test design, development, and evaluation.

The intention is not to provide a definitive guide to testing. But more to explore the current state and future of test development, opening a window into what our expert team of psychometricians are currently developing and researching. This guide shares some of the advice that our team is giving to clients on a daily basis, structured across 3 key stages involved in test construction: design, development, and evaluation.

What is psychometrics?

Psychometrics, or psychological measurement, is a scientific discipline concerned with the construction and evaluation of assessments that measure knowledge, skills, abilities, traits, and attitudes. All of these techniques help organizations to connect observable phenomena to underlying attributes.

What does a psychometrician do?

A psychometrician is someone who practices the science of educational and psychological measurement, or in other words, testing. A psychometrician constructs tests and interprets results for the purpose of assessing a person's psychological attributes.

Contents

02

Test *design*

The landscape of test design has changed dramatically over the last few years. Testing capabilities have increased exponentially with advances in computing power, and the introduction of game-changing technologies. This section of our guide covers some of the key learnings that have emerged when it comes to test design.

"The landscape of learning has already changed dramatically over the last few years. We need to be constantly adapting assessment to reflect contemporary practices and theories of learning."

Applying learning sciences to assessment design

How compatible are learning sciences and psychometrics? Both fields make inferences about learner cognition based on manifested behaviours. However, they diverge in how they conceptualize the design and use of assessments. What are the opportunities and challenges in bridging these two paradigms, and what does this mean for the future of assessment?

Psychometricians seek reliable ways of measuring knowledge in well-defined domains. Accordingly, assessments should be decontextualized to minimize confounding variables and maximize measurement reliability. This gives test takers equitable opportunities to demonstrate their knowledge and skills, regardless of context or test form.

In learning sciences, a central challenge has been designing assessments in line with evolving theories of learning. For example, constructivist theory posits learning as an active, context-dependent process of constructing knowledge based on prior understandings. Learning scientists are therefore concerned about the degree to which test takers transfer knowledge to new situations by leveraging prior knowledge, context, tools, and other people in their learning environments. In this view, assessments should measure deep conceptual understanding and complex constructs that are predictive of success in the workplace, like capacity for scientific inquiry and collaboration. Ideally, assessments would be a part of the learning process, and the evidence would be knowledge traces extracted from video, audio, and log data.

The differences between the learning sciences and psychometrics stance toward assessments can be boiled down to four issues: what to measure, what counts as evidence, what inferences we can make based on this evidence, and what to do with this evidence.

Underpinning these issues lies a fundamental divergence in how to view validity: learning scientists value complexity and ecological validity, whereas psychometricians value reliability and construct relevance.

Advances in computing power and artificial intelligence have the potential to bridge the two paradigms. Moreover, the digital world now affords a medium where we can observe many behaviours in a realistic setting. For example, researchers are using log files to measure more complex constructs such as collaborative problem solving and persistence. Such constructs are indicative of success in the workplace but have traditionally been difficult to measure. More data can help address issues of reliability and create a more nuanced understanding of learner knowledge.

Rethinking traditional notions of validity for more innovative assessments may seem pie in the sky and not worth the effort. However, we need to be constantly adapting assessment to reflect contemporary practices and theories of learning. Already the landscape of learning has changed dramatically over the last few years. It's difficult to say what kind of game-changing technologies five years from now will change the way we live, work, and learn. As psychometricians and learning scientists, what we can say is that it's an exciting time to be part of such a dynamic field and see the changes in store for assessments in the near future.

How many items are needed to create an exam: *A blueprint for success*

If examination items are the building materials used to construct a test, then an Examination Specifications Document is the blueprint for building a test. In any building, materials need to be repaired or replaced due to wear-and-tear, unexpected damage, and changes to building standards and regulatory requirements. Likewise, examination items have a shelf life that is affected by regular exposure to examinees, exam security incidents, and changes in the practice area being assessed. It follows then that two of the pressing questions in designing an exam development process are: How many items do we need, and how often do we create more?

At bare minimum, an examination program requires enough items to be able to publish an examination that conforms to the Examination Specifications Document. If we have told test takers and other stakeholders that there will be 80 scored items and 20 unscored items and that there will be a set number of items allocated to certain content areas, then we are bound by that.

That said, anyone who has worked with knowledge-focused examinations will tell you that items on these exams are not evergreen. There is always an attrition rate whenever we evaluate items, and that is because statistical analysis and content review can help us identify content-related and performance-related issues with items. As such, ensuring that we have 100 items today but no more leaves us in a precarious position for the future unless additional item development is underway.

There will always be a need for more exam items, but how many more?

It all comes down to item usage, which is sometimes envisioned as exposure (or the number of eyes looking over your items). Imagine that we expect 1500 examinees to take our 100-item exam in the coming year. It would be folly to have only one form of this exam available at a time because we can anticipate that items will be increasingly discussed by examinees over time and there is always the risk of security incidents that could compromise content.

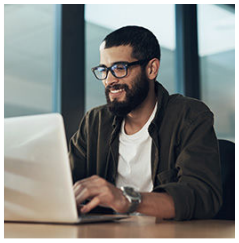
In this case, we know that we need more than one form at a time, but the exact number of forms is determined by the amount of item exposure that can be tolerated by the certification program and must be balanced with the resources available for exam development. If we have a low risk appetite, or suspect that cheating is more likely to occur, we may opt for four forms, which caps the number of item views at 375. If fewer resources (e.g., SMEs, pounds, pretesting capability) are available, or we have a higher risk appetite, we may opt for two forms because we feel comfortable with 750 item views in the coming year.

There is a flipside to this: If we have another exam that has 50 examinees in a year, then more than one form would be ill-advised because there is little utility to limit exposure even further – and it would decrease the interpretability of statistical analysis if the data we receive is based on 25 examinees per item rather than 50. Our quest to improve exam security shouldn't introduce other serious issues, like score reliability.

It should be noted that an increasing number of credentialing programs look to pool-based approaches like linear on-the-fly testing (LOFT) rather than using static fixed forms because it allows for dynamic updates to examinations. In any case, the same line of thinking allows us to determine the size of the item pool: If we have 1500 test takers and we want to ensure that items are not viewed more than 500 times, then we would look to create a pool size of 300 (or 3x, as in three times the size of what is called for by the examination specifications).

The frequency of exam development depends on the expected rate of change to examination content. For example, a program that assesses technology-focused competencies may opt for exam development to occur every six months because changes in practice occur very frequently, while a program that assesses interpersonal competencies may opt for exam development to occur every two years because changes in practice occur infrequently or very gradually. In either case, we need to ensure that we have enough forms, or a large enough item pool, to accommodate the length of time in between publications.

In our metaphorical building, it would be fair to expect that areas more frequently used (like the kitchen or living room) will need more frequent upkeep than less-used areas (like the basement and attic). We know we can't take our responsibility to examinees and just brush it under the carpet, so we continue to rebuild, repair, and refresh our metaphorical home for examinees. That said, tidy and well-behaved house guests leave less work for us to do than guests who outstay their welcome. And who can forget that there is only so much money that we can spend? Exam items don't grow on trees after all!



"*Examination items have a shelf life* that is affected by regular exposure to examinees, exam security incidents, and changes in the practice area being assessed. It follows then, that two of the pressing questions in designing an exam are: How many items do we need and how often do we create more?"

Making the most of a job analysis without in-person interaction

Various organizations and industries have adjusted their work processes due to COVID-19. Call it what you may – adapting, altering, accommodating – the acknowledgement that process changes may be needed have helped these businesses face these inherent challenges head-on.

The job analysis, often cited as the building block of a certification program, has not been immune to these challenges. As travel and large group gatherings have been limited, the virtual job analysis has gained in popularity. Below we outline how to make the most out of a job analysis without in person interaction.

While it is true that our current circumstances have played a large role in the increased interest in virtual job analyzes, it is not necessarily the only impetus. The advantages of in-person meetings, which admittedly should always be considered, are counterbalanced by the inherent limitations that they pose. Some of these limitations include the fact that:

- Not every job analysis decision can be made all at once in a meeting.
- In-person meetings may inadvertently exclude certain perspectives.
- In-person meetings are expensive.

It is the combination of these two forces, COVID-19 and the limitations of in-person meetings, that has led to a rise in virtual job analysis. Now, let's take a look at the steps that can be taken to run the most efficient and productive virtual job analysis meetings. Utilising these steps helps maximize any virtual

meeting. When considering whether to conduct a job analysis virtually or in-person, it is important to consider the pros and cons of each method. While a virtual job analysis may offer decreased costs for participant involvement, as well as greater flexibility in terms of scheduling, it may also provide less incentive for participation without the added bonus of an in-person trip.

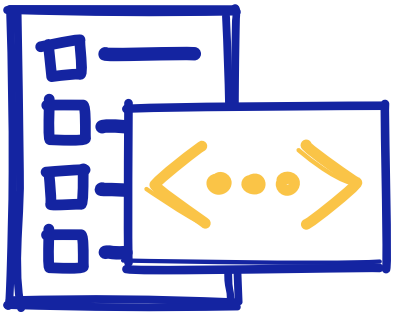
Fatigue and technology challenges must also be considered. Either way, considering the environment and challenges that are posed by either an in-person or virtual job analysis before the actual meetings can go a long way to ensure a smooth experience for both you and your subject matter experts.

- **Communicate early and often –** It's important to outline meeting objectives to participants beforehand. This helps eliminate any ambiguity that may be present on the day of the meeting.
- **Troubleshoot technology –** The software tools that help make virtual meetings possible are great, but they are not bulletproof. Provide participants with the opportunity to test out the technology that will be used before the meeting.
- **Set expectations –** Make sure that participants know that, even though the meeting is not in-person, all interactions should still be respectful of the input and contributions of others.
- **Take advantage of smaller group sizes and meeting times –** Leverage the virtual format by getting creative with the participant groups and meeting times. You may have the opportunity to meet with smaller groups, and meeting times can be split (e.g., two 4-hour meetings instead of one 8-hour meeting).

Top tip

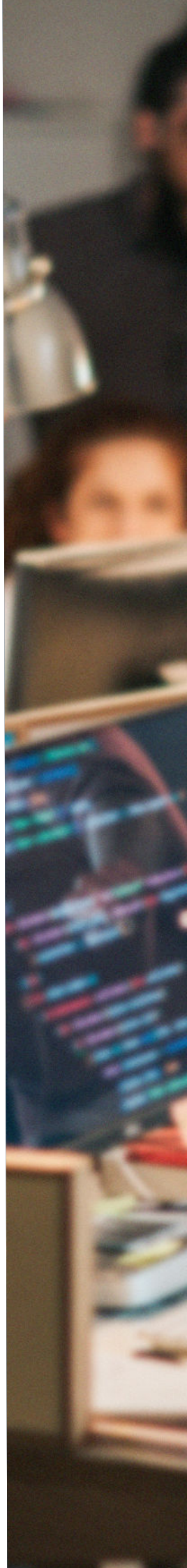
Considering the environment and challenges that are posed by a job analysis before the actual meetings – whether they are in-person or virtual – goes a long way to ensuring a smooth experience. For both you and your subject matter experts.

03



Test *development*

Your tests need to be fair. Should it come to it, you always need to be able to defend your decisions in court. Test content must also be valid. The questions you ask should reflect the attributes that are relevant to your program, and the job role that you are assessing competency in. This section of our guide looks at some of the ways you can ensure that your tests are fair, defensible, and valid.



Referencing items for certification programs

Certifying bodies vary in how they approach referencing. Some require every item on an examination to be referenced, some identify a reference for items in most instances, and yet others rely on the consensus judgement of a group of subject-matter experts (SMEs). Although some items may be difficult to reference, our recommendation to clients is for all examination items to have a reference.

Why should we reference every item?

The main reason we recommend referencing is to support the validity of the examination content. Certifying bodies often receive emails and phone calls from test takers challenging examination content and thereby the outcomes of their testing experience. Although it may be tempting to respond that a group of SMEs agreed on the correct answer, pointing to an authoritative reference in the test taker's profession that corroborates the SME judgements carries much more weight.

Another reason we recommend referencing every examination item is to ensure the defensibility of examination results. In very rare circumstances, this means defending an examination in court. In such cases, certification examinations become more defensible when SME judgements are combined with authoritative references.

What are some recommendations for item referencing?

- **Have an approved reference list.** Test takers should be able to use a list of approved references to prepare for their certification examination. However, reviewing those references should not be required in order to pass the examination. Certifying bodies should strive for a concise list of approved references, yet with a sufficient number to cover the entirety of the examination's scope of content.

- **Have a policy for referencing items.** The certifying body should establish general guidance on acceptable item referencing practices. Such a policy may address which types of references are acceptable, how often references should be updated, how many references are needed per item, and so on. Textbooks, journal articles, and position papers are the types of references most often used in certification. Certifying bodies should consult the examination development committee or other SMEs to establish the referencing policy and how frequently it should be updated.

- **Incorporate referencing into item writing.** An item should be referenced when it is initially written, and that reference should be confirmed throughout the item development process. When items without references are selected during examination assembly, the examination development committee must spend time referencing items during their review of the examination, rather than focusing on examination content. This is an inefficient use of the committee's time and makes the review process longer than necessary.

- **More is sometimes better.** It is easy for SMEs to reference facts and definitions straight out of a textbook. However, we often find SMEs struggle with content that requires application of knowledge, such as clinical judgement or evaluation of a novel issue or situation. Nevertheless, identifying a reference that can support practitioner judgement is helpful, even when the reference does not describe the exact situation. In cases where an item is more complex or requires the respondent to evaluate competing sets of information, it may be necessary to have references for each principle required to arrive at the correct answer.

We want to ensure each item that appears on an examination form is fair, accurate, meaningful, and current. Enforcing a policy on referencing items is an excellent method to ensure the content of each item is factually accurate and up-to-date. Item referencing also makes examination results more defensible and the examination development process more efficient. So given the benefits, why wouldn't you reference items?

Top tip

Enforcing a policy on referencing items is an excellent method to ensure the content of each item is factually accurate and up-to-date. It also makes examination results more defensible and the examination development process more efficient.

A deep dive Into *cognitive Levels* and a case for simplification

If this is your first time hearing about the concept of human cognition as it relates to certification examination items, you're not alone. It's not a particularly high-priority subject among psychometricians because, after all, we cannot calculate an index on it.

Inevitably, it cannot be a subject we ignore either. The concept of an item's cognitive level, or thought process required to arrive at the correct answer, is an integral piece to developing a high-quality certification exam.

The notion of item cognitive complexity is derived from a 1956 publication by educational psychologist Benjamin Bloom, The Taxonomy of Educational Objectives: The Classification of Educational Goals. Dr. Bloom was discouraged by the state of educational tests with regard to the preponderance of items testing lower-level thinking skills. His intent was to persuade writers for educational objectives to write items that engage different levels of cognition, in particular, higher-order thinking skills. His publication outlined a rubric for six levels of increasing cognition objectives including knowledge, comprehension, application, analysis, synthesis, and evaluation. In other words, the main intent of the rubric was to ensure that not all educational objective items are written to elicit recall of facts; rather, that they require some additional amount of thought in order to respond.

Notice that in the above paragraph that the word "education" is heavily emphasized. Suffice it to say, the educational and certification landscapes are different in scope and outcome. Whereas in the educational landscape, a psychometrician is interested in classifying students into a multi-level tier of proficiency (e.g., basic, intermediate, advanced), psychometricians within the certification landscape are interested in whether a test taker fits into one of two classification levels: competent or not. So how does a rubric developed with educational objectives in mind translate to the world of certification exams?

The point of cognitive levels in credentialing is to ensure exam content validity. An exam that reflects the complexity of situations lends one piece of content validity evidence. That said, not all items on an exam need to be complex because many important tasks in any job may not be complex. So, as a professional working with credentialing exams, how prescriptive do we need to be about cognitive levels? Not all certification programs are built the same, and therefore, there is wide variation on how cognitive levels are addressed and adopted within certification programs. Some programs use Bloom's taxonomy verbatim, others often rework (and often condense) the levels into different groupings inspired by Bloom (e.g., recall, application, and analysis), and some don't address cognitive levels at all.

All that said, in our experience, almost every item can fall into one of two categories, recall and not recall. It doesn't matter how "not recall" is termed (e.g., application, analysis, evaluation, synthesis). Whatever it is, we can be comfortable knowing that this item is measuring something other than information memorized from a textbook. "Not recall" items go beyond whether an examinee knows something; instead, these items venture into the territory of whether an examinee is able to do something.

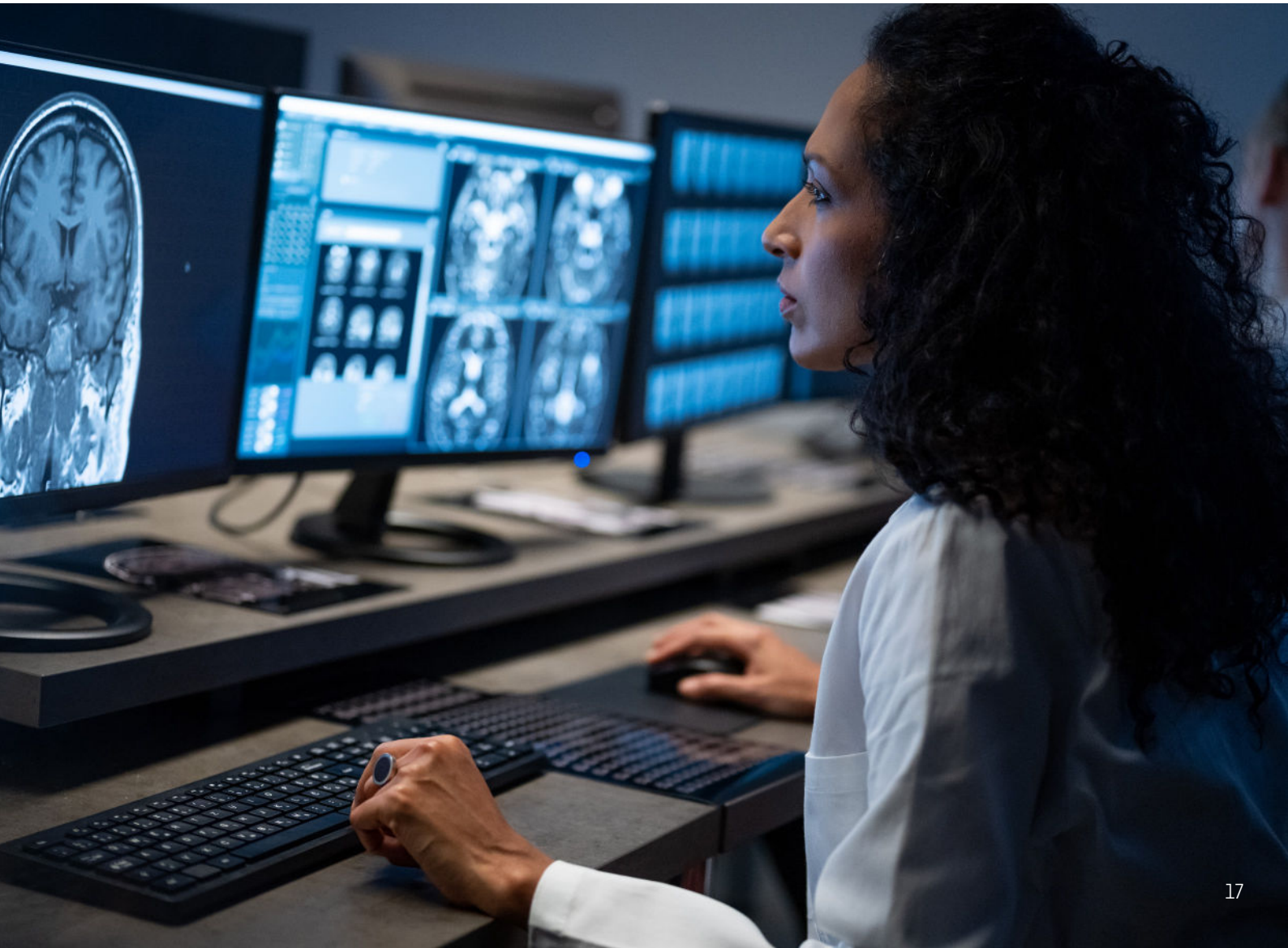
It is our experience when exam committees use a three-tiered cognitive classification scheme, they are often faced with some discomfort when making a final determination of the item's cognitive complexity. When asked to verify the cognitive complexity of a "not recall" exam item in a three-tier rubric, we often hear "It's on the fence. Where do we need it more?" Theoretically, this is because there is considerable overlap, particularly at the two higher levels. In these times of unease, exam committees force items into a level for superficial reasons. This situation highlights the subjective nature of the rubric and works against our goal of making the world a more objective one.

Whether cognitive levels are used as part of your

exam blueprint or not, it is the responsibility of a psychometrician to educate item writers and reviewers regarding different levels of cognition. After all, this training will prevent us from developing an exam that consists exclusively of recall-type items – which is as Dr. Bloom intended. After training, however, the level of complexity on an exam should belong in the hands of the exam committee, with guidance from the psychometrician. The cognitive level requirements for exam items as a requirement of a blueprint should not place undue burden for developing the exam. Exam committees should not be forced to classify items based on need. Alleviating these restrictions with a simplified two-tiered classification (or having no cognitive classification system at all) may be the best approach with certification exams to ensure that the measurement is sound and job-related.

Top tip

A simple two-tiered classification – or having no cognitive classification system at all – may be the best approach with certification exams to ensure that the measurement is sound and job-related, and not overly focused on what may end up being an arbitrary judgement on the mental process required to answer an item.



Subject matter experts: The key to your credentialing examination

One of the first steps in any content analysis or form development process is to recruit subject matter experts (SMEs). Ensuring that the right perspectives are counted is critical to the content validity of a knowledge-referenced assessment. But who are your SMEs? What characteristics lend themselves to a quality SME? How many SMEs should there be?

Step 1: Define the target population

The initial step of determining a SME is determining the target population. The target population is the "pool" of individuals that SMEs should be pulled from. The target population should consist of individuals who currently hold the target credential, currently hold a related credential with similar eligibility requirements, or are currently eligible to hold the credential. Furthermore, these individuals should have recent (e.g., within the past 6 to 12 months) work experience in the profession. These individuals are best suited to providing insights into the target job role or areas of knowledge related to a skill set.

Every target population can have different demographic and professional characteristics of interest, such as: years of experience, level of education, geographic location, specialty. Every collection of SMEs (whether that is committee, panel, task force, or even just a handful of individual contributors) should be a reflective sample of the target population's demographic characteristics. For example, if 25% of the target population live and work in London, roughly 25% of the SME committee should be comprized of individuals who live and work in London. It is fine if the committee's demographics are not a perfect match compared to the target population as long as it is similar to or reflective of the target population.

Step 2: Recruitment

Research suggests that 12 to 15 individuals or 10 to 20% of the target population, whichever is most practical, allows for sufficient diversity regarding the demographic characteristics described previously. In any case, it is preferable to target more potential SMEs than necessary due to the likelihood of attrition. While potential SMEs may be eager to be part of the process, some may not realize the amount of time that is truly required of them and may resign midway through – or worse, do nothing. Attrition with an already small SME committee could be detrimental to the progression of the project and possibly push back the implementation of a new form. A common tactic to entice SMEs to serve on these committees is to present an incentive, such as a few continuing education credits towards certification renewal or nominal honororium, in return for their service.

A question that may present itself is: Who should I look for? Treat the recruitment of SMEs quite similarly to the recruitment of a normal paid position. Request cover letters, resumes, and/or curriculum vitae, and then interview those who appear promising. It is best to get a mix of familiar faces and perspectives. Recruiting SMEs who have previous experience with exam development can be a bonus, but representativeness of the group can be skewed if first-time contributors are not also included.

Step 3: Get to work

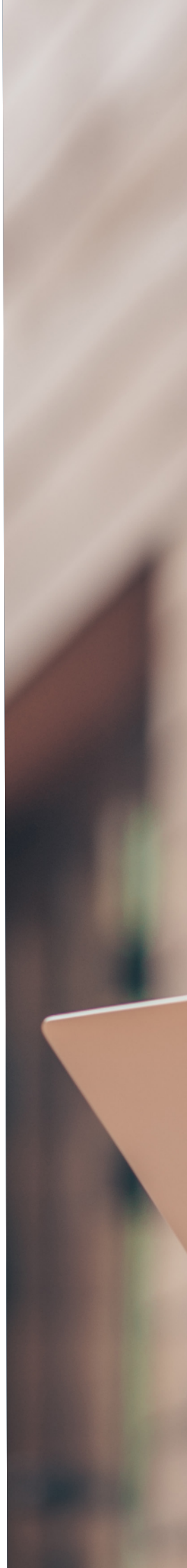
Once you have defined the target population and assembled your committee, it's time to put them to work and create your credentialing examination.

"Ensuring that the right perspectives are counted is critical to the content validity of a knowledge-referenced assessment. But who should you involve and how many SMEs do you need?"

04

Test *evaluation*

We have a sizable amount of statistical knowledge and expertise on the team here at PSI. Our aim is to always apply this knowledge to the real-life questions and decisions that our clients are facing. This section of our guide looks at some of the big questions we are most frequently asked when it comes to test evaluation.



"By asking a specific question in a particular context, you'll drastically reduce the number of factors that could affect the problem at hand. Therefore reducing the likelihood of an "It depends" answer and helping us (your psychometrician) formulate a more detailed and useful response."

Getting beyond "It depends": How to get the best response from your psychometrician

We've seen some funny quips about questions, answers, and the nature of truth in measurement. For example: "A man with two watches never really knows what time it is." Or this one: "Ask two psychometricians a question, and you'll get at least three opinions." These quips are funny because they are true. A common response from psychometricians is also true, but some may not find it funny anymore! We're sure it's happened to you: you ask a psychometrician a simple question, and the answer is "it depends."

You: "What's the ideal sample size for equating?"

Psychometrician: "It depends."

You: "What measurement model is best?"

Psychometrician: "It depends."

You: "Is the Bookmark Method of standard setting useful?"

Psychometrician: "It depends."

While in conversation with a colleague, she broached the subject and suggested we put a moratorium on "it depends." Surprisingly, she was speaking as a psychometrician about "we" as a community of psychometricians, showing that even some psychometricians have grown a little weary of this response! The problem with "it depends," though, is that it is true. The answers to the questions above DO depend. Because when faced with questions related to assessment or measurement, lots of factors need to be considered: technical limitations, business limitations, availability of resources, time, goals, risks, benefits…the list can go on and on.

Getting past "it depends"

If we are going to get past "It depends" as the first answer to so many questions, both psychometricians and non-psychometricians can play a role. For psychometricians, answering with "it depends" is too easy and may seem to others like we're trying to avoid the real discussion - or even worse, that we're being lazy with our answers. So, as a psychometrician, try to not even say "It depends" until you're absolutely ready to talk about what it depends on!

Non-psychometricians should beat us to the punch. Try to ask questions in a particular context. This will drastically reduce the temptation for us psychometricians to say "It depends." Below are suggestions for better ways to phrase the previous questions:

- "Given our program's history, how many test takers do we need to have confidence in our equating results?"
- "What measurement model makes the most sense for our program?"
- "Do you think the Bookmark Method would be useful to set the standard after our next job analysis?"

By asking a specific question in a particular context, you'll drastically reduce the number of factors that could affect the problem at hand, therefore reducing the likelihood of an "It depends" answer and helping us (your psychometrician) formulate a more detailed and useful response. It can be useful to know what other programs might do, but it's even more useful to understand what solution paths are more or less promising for you.

In those inevitable situations where you still get "it depends" as an answer, try to be patient with your psychometrician. Perhaps simply nod encouragingly, and the rest of the answer will unroll with a collection of factors and points of view to consider. And remember, psychometricians often resort to this overused phrase because one size typically does not fit all and there are few absolutes when it comes to measurement.

Scaled Scoring for your certification examination

You've probably heard of score scaling and may have wondered how it works and what purpose it serves. The process of scaling is simpler than you may think, and understanding it further might help you determine whether it can be helpful for your program.

What is scaling?

Often times for certification examinations, scores are initially computed on a **raw scale** (sometimes called the true score scale), where a total score is just the number of items a test taker answered correctly, and each point represents one correct answer. Colloquially, we use the term **scaling** (or "scaled scoring") in reference to when raw scores are converted to a different scale, typically for reporting scores to test takers. So, technically, scaled scoring is a different kind of scoring process – it is a different way of presenting scores.

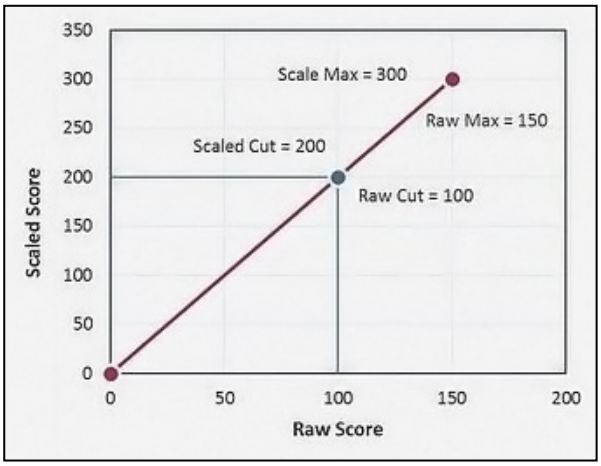


Figure 1. Scaling Example

In the example of scaling in Figure 1, a 150-item examination with a raw cut score of 100 is converted to a scale of 0 to 300 with a scaled cut score of 200. Scaling does not change the length, difficulty, or number of items answered correctly required to pass the examination. Similar to converting from inches to centimetres, scaling just changes the scale on which scores are reported.

When is scaling useful In scoring your certification examination?

Any examination can be scaled, but some circumstances may make scaling a more attractive option over reporting raw scores. Here are a few situations where scaling may be preferable:

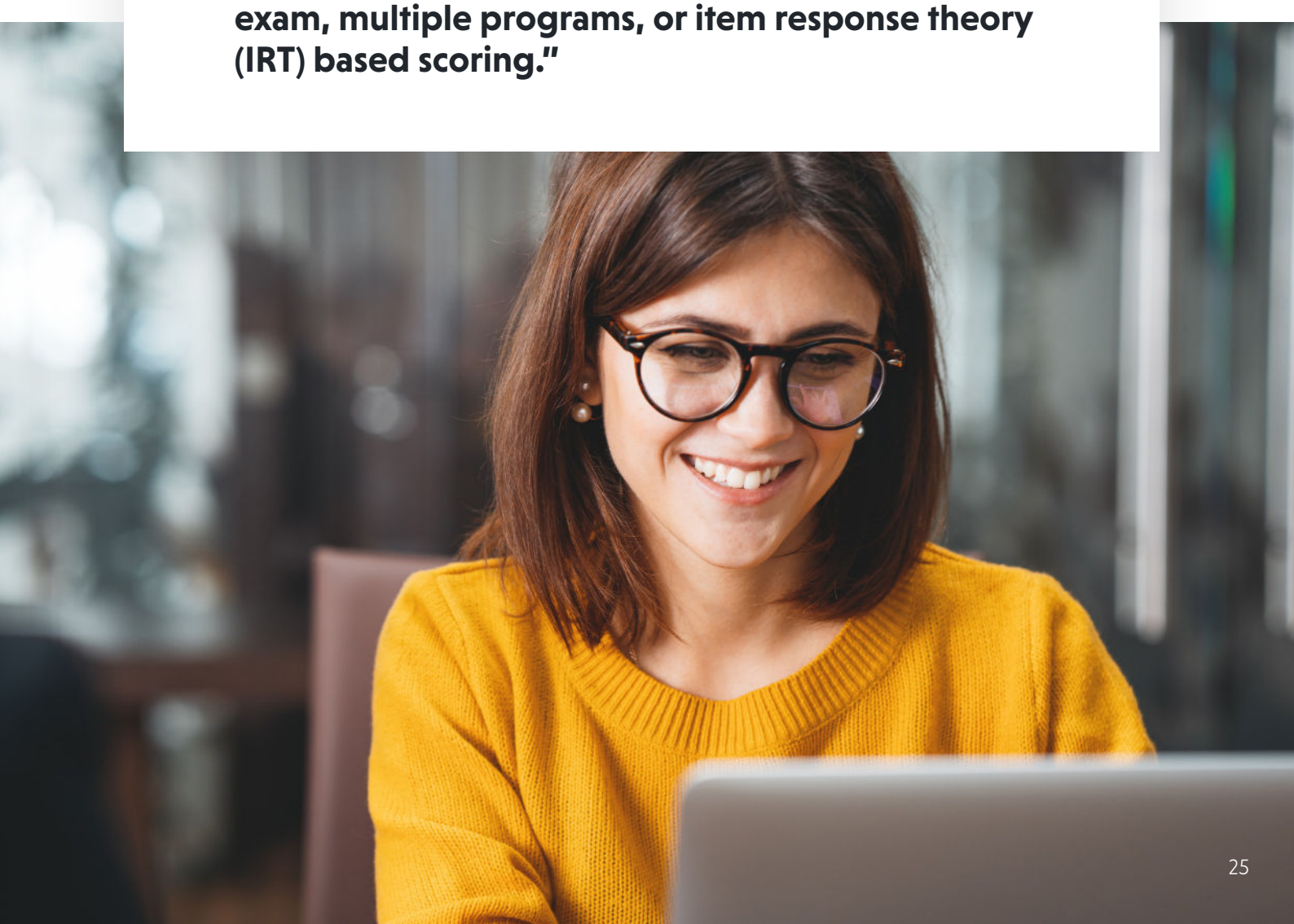
- When the raw cut score changes from form to form. This may happen if two assembled forms could not be made parallel in terms of difficulty, or if equating is conducted after an administration window. By placing all forms of an examination on the same scale, the credentialing organization can provide a consistent means of score interpretation. Furthermore, different cut scores for different forms of an exam can sometimes confuse test takers (e.g., two test takers get the same raw score on different forms, but one test taker passes and one fails). Though outcomes may be psychometrically and statistically justified, they can have the appearance of being unfair; scaling can help mitigate such confusion.

- When the credentialing body offers several credentials and examinations. Many credential-granting organizations have more than one credential, and often the examination lengths and cut scores differ across programs. Placing examination scores for all programs on the same scale may present a more cohesive face of the programs to test takers. This may also allow for more consistent explanations of scoring and scaling in the respective test taker handbooks.

When the examination scores are not based on number of correct answers. This can be the case when employing Item Response Theory (IRT) for scoring. Without getting too much into the detail, some IRT models provide an estimate of test taker ability on a scale that is not terribly conducive to meaningful test taker interpretation (e.g., -3 logits to +3 logits). In the case of computer adaptive testing (CAT), scaling the ability estimates is a necessity because different test takers receive a different number of items to determine their score.

Scaling can be a fairly straight forward conversion of raw scores, and can help establish consistency of score reporting. In providing that consistency, scaling allows for better management of the score interpretation, which may be desirable for credentialing organizations with multiple forms of an exam, multiple programs, or IRT-based scoring.

"Scaling can *help establish consistency of score reporting and allow for better management of the score interpretation.* This may make scaled scoring desirable for organizations with multiple forms of an exam, multiple programs, or item response theory (IRT) based scoring."



05

In *conclusion*

We have a team of incredibly talented psychometricians here at PSI. Many of our clients tell us that's a big reason they love working with PSI. What's more, our team of psychometricians aren't just brilliant, they're also accessible. We don't hide our technical experts away. It's part of the PSI culture.

Our team of quality psychometricians are heavily involved with client projects from day one, providing knowledgeable guidance and advice to our clients on a daily basis. Come and test our psychometricians.

Your trusted *exam partner*

Every day our clients support millions of people to realize their dreams, reach their potential, and improve their life chances. They care about their test takers – and we share that responsibility.

Our unwavering focus is on maintaining exam security and integrity, while delivering frictionless and fair test taker experiences, through...

Secure exam delivery

- Authorized global **exam center** network.
- Secure and scalable **remote exams** with Live online proctoring.
- Flexible **multi-modal** exam delivery.
- Testing **windows or continuous testing**.

Rigorous exam development

- Legally defensible and **valid exam content**.
- Exam reviews to keep pace with regulatory and industry change.
- **Job analysis** and **exam** content specification.
- **Subject Matter Expert** (SME) recruitment, training and management.
- Secure **item authoring, banking and exam generation** software.

Expertise in testing science

- Experienced **psychometricians.**
- Specialist **exam developers**.
- **Data forensics** and web crawling.

Our willingness to listen and adapt means clients can either benefit from a full service, or access solutions at any stage of their testing journey.

Dreams deserve

We understand every exam is about more than the result. It's about a dream. A dream the test taker believes is worth striving for. And we believe that too. Their dreams deserve trusted science, technology and operational expertise. They deserve PSI.

Connect with an expert today.

psiexams.com