CU-101.6_Intro to Item Writing_Alan Mead_for review

0:01
Welcome to Certiverse University and welcome to this video introduction to item writing.

0:05
My name is Alan Mead and I'm the chief psychometrician for Certiverse.

0:09
In this video, we'll cover the essentials of leading a team of subject matter experts, which I'll call SMEs, in writing effective exam items.

0:16
Our learning objectives cover all aspects of effective item writing, from planning item assignments all the way to item review, so that you can achieve your goal of ensuring high quality, fair exams.

0:27
I'll also share some research on best practices using AI to assist me's at writing items.

0:32
Now that's a lot of content, but by the end of the video, you should be prepared to lead a team of item writers.

0:36
Let's get started.

0:39
Item writing is crucial for creating fair, reliable, and valid certification exams.

0:44
In the next video, we'll talk about beta testing items and how to identify poorly performing items so they can be removed.

0:50
That's great, but at that point, we can't fix those items.

0:53
We can only remove them.

0:54
That's why effective item writing is absolutely critical and also why you should write slightly more items than you want to finish with, maybe 10 to 50% more.

1:04
But don't be intimidated.

1:05
We'll cover best practices that will help ensure the best possible items.

1:11
Now the first step in item writing is to create item writing assignments, which involves a bit of math.

1:16
And as a decide, I want to share with you that I myself not a big fan of math for math sake.

1:21
And if you also don't love math, then here's my advice that I give my students.

1:25
Just ignore the math and instead try and extract the qualitative point like here with planning item assignments.

1:31
The point is that you need a good mix of assignments that align with the weights of the blueprint.

1:36
If you're a Certiverse user, by the way, you won't ever have to do these calculations because the system does them automatically behind the scenes.

1:41
So if you're not a fan, don't let the math dissuade you from getting the main point.

1:45
Now about the item assignments.

1:46
They should be proportional to the requirements.

1:49
If you wanted 55 scored items for two forms and a 20% margin, then you need about 132 items.

1:57
If the blueprint weight for a topic 1.1 is 15%, that's about 20 item assignments.

2:04
I usually let SMEs decide which 20 items they want to write within that topic or subtop.

2:10
If 20 seems like too much latitude, and that's an indication that in the previous step, a more detailed blueprint would have been better, I wanted to say a few words about recruiting subject matter experts, whom I call SMEs.

2:22
SMEs possess the necessary technical expertise in the content area of the exam, so they will write and review all items.

2:29
An item author is a Smee.

2:31
Bonafide.

2:32
SMEs are often current or shammer role incumbents, that is, they work or have worked in that role, but sometimes they are closely associated with that role, for example, the supervisor of that role or they teach the subject.

2:47
Now, methods to recruit qualified SMEs include seeking recommendations from people on your team, using professional networks and advertising on freelance sites.

2:57
It is very common, but not ubiquitous to offer incentives.

3:03
But that could have its own pitfalls and you need to weigh what's best for your exam program.

3:07
Whatever you do, you should try hard to recruit SMEs that reflect the diversity of your candidate population in terms of demographics, industry, role, tenure, etcetera.

3:22
A question that I get a lot is how many SMEs do I need?

3:25
And one answer is that more is usually better.

3:27
But let's do the math.

3:29
If you assume one hour per item, and we want 132 items, that's 132 hours of effort.

3:37
Then we need to make an estimate of how many hours a Smee spends per week.

3:40
Let's say it's four hours per week.

3:42
In that case, with two SMEs it will take 16.5 weeks, almost four months.

3:48
With five SMEs it will take 6.6 weeks, and with 10 SMEs it will take 3.3 weeks.

3:54
But usually some SMEs are more active than others, which I think is true of work in general.

3:59
That SME productivity follows a Pareto distribution, just like performance on other tasks.

4:05
So the more the better if you can manage it.

4:08
And also, everyone's excited about AI today, and I think with good reason.

4:11
I have data from a very small sample that suggests that using AI is associated with a 60% reduction in time spent writing items.

4:18
And that's either very good news or very bad news.

4:21
If you thought AI would do all the work, then I guess that's bad news, but a 60% reduction is still pretty good.

4:28
This is a very high level overview of guidelines for writing items.

4:32
First, items should be written to a specific target domain or sub domain as specified by the item assignment for that item.

4:39
Question stems.

4:39
Incorrect options should be clear and concise.

4:42
The stem is the part of the item before the response options that sets up the question.

4:47
It might have a scenario or some context and usually ends in a question.

4:51
It should also be the biggest part of the item.

4:53
Now obviously the correct answer must also be demonstrably correct.

4:57
Distractors are the incorrect options, and they should be plausible and yet demonstrably incorrect.

5:04
You should avoid common pitfalls like ambiguity and tricky questions.

5:08
And finally, you should follow all best practices for writing high quality items.

5:13
Now, if you wanted more detail, in 1989, Haladina and Downing published 43 rules for writing multiple choice items, and these rules have become very widely used.

5:22
If you search for rules for writing multiple choice items, the first several pages of hits are mostly variations of these rules.

5:29
And there are 7 rules about general procedures such as #5 Use good grammar, spelling, and punctuation.

5:35
There are 10 general guidelines about exam content such as #15.

5:39
Avoid items based on opinion.

5:41
There are 6 guidelines about SEM construction, for example, rule #21 Avoid unnecessary verbiage.

5:47
And finally, reflecting how critical the response options are, there are 12 rules about them, such as #28.

5:53
Keep option links fairly consistent.

5:56
You should become familiar with these 43 rules that I have included the reference, but they are a more detailed elaboration of the overview on the last slide.

6:06
So are these rules or guidelines?

6:08
Are they based on empirical evidence or opinion?

6:10
Well, these are guidelines and they're based on wisdom rather than evidence.

6:16
So for example, one guideline is to phrase item stems positively, and this can be tricky to even understand.

6:23
Are you forbidden to write items about nonprofits because it has non in it, or not for profits because it is not in it?

6:29
If you're writing an IT exam, you can't say a web server is not responding because it has not in it.

6:35
No, that would be silly.

6:37
The key issue is clarity.

6:39
There are questions that can be best expressed and perfectly clear with negative phrasing, and clarity is paramount.

6:46
But what you need to be vigilant about is creating double negatives that can be very hard to understand and that are likely to create trick questions.

6:54
Also, evidence I've seen is that negation has a very slight effect.

6:58
Items using negation are slightly less likely to survive beta testing, but are as good or slightly better than non negated items on operational forms.

7:09

So these are guidelines that you have to apply judiciously to the item writing process.

7:16

Now I'd like to share some research with you that we've conducted on using AI to assist Smeeze and writing items.

7:23

We compare different AI models by asking them to generate items from the same specific topics for a psychometrics exam blueprint.

7:31

We as Smeeze then rated the items on a one to four scale where one was useless, 2 needs substantial edits, 3 needs minor edits, and four is ready to go.

7:40

And you can read our paper and online at the Journal of Applied Testing Technology.

7:45

I've included the reference for you at the bottom.

7:48

We review the recent literature on generating items.

7:50

We talk about how large language models work, talk about our methodology, and you can see all the items written by the models Now.

7:57

Our main findings are that AI models had a high rate of successfully writing usable items, with 71 to 90% of the items deemed usable by subject matter experts.

8:09

Although many items are useful, the typical item will require some revisions to correct issues with the question stem, correct answer, or incorrect answers.

8:17

The most frequent error was violating the standards of multiple choice question design, such as including more than one correct answer.

8:24

In this study, we compared two versions of Jeep D3, but since then we've compared Jeep D4 and Gemini.

8:30

Newer models are slightly better, rising to a modal rating of three out of four, which means the average item will need some edits.

8:40
So these models will not replace item authors, but they show great promise in assisting item writers by automating parts of the exam creation process, which saves time and resources for exam developers.

8:52
So we still need human SMEs, and we select SMEs because they're experts, but not experts at writing items, which is a unique art.

9:00
Therefore, we typically train item authors, and the most traditional training has two parts.

9:05
1st, there's a list of house rules about the items.

9:08
For example, if that is about legislation, it would include standard ways to refer to legislation, or rules about referring to people by last name to avoid queuing gender.

9:17
The second part is a long list of better and worse examples.

9:20
Based on Haliden and Downing's 1989 list.

9:23
You might not have an example of all 43 rules, but this is typically a long list to lead the authors through during the training.

9:30
But, and this surprises me a lot, we have 0 research studies comparing training methods for item authors, and if we apply lessons from performance appraisal, which is a parallel literature, then training SMEs with long lists of do's and don'ts is less effective than training on examples.

9:49
So applying this to item writing, I think that practice writing items is quite important.

9:54
The handbook chapter on this topic emphasizes the need to give feedback to item authors as part of this training.

10:03
I suspect that it's also helpful to serve as a peer reviewer and seeing mistakes that other authors make and making suggestions to correct them.

10:12
And finally, because it's a learning curve, you should recognize the value of seasoned item writers when recruiting SMEs.

10:18
Getting SMEs to the top of that learning curve quickly may be another benefit of using AI to assist the item writers.

10:25
The peer review process addresses best practices so the topics closely aligned with item writing best practices.

10:32
Peers should verify that all required elements are present, such as the stem and the key and options, difficulty rating, authorship, justification, references.

10:40
They should ensure correct formatting.

10:42
Of course, online systems try to automate this as much as possible.

10:46
The item should be clear, properly worded, non offensive, and free of unnecessary language.

10:51
Peer review should also confirms the item aligns with the assigned content, focus on a single topic and uses appropriate terminology.

10:58
It's important that the stem provide all necessary information, asks a clear question, and avoid assumptions.

11:05
It is obviously extremely important that reviewers check that the key is correct, that options are independent, distractors are plausible, and ensure that no clues are yet.

11:15
And finally, there are other requirements that might be specific to your exam program.

11:18
For example, the Certiverse platform requires rationales and references that the SPE must evaluate as part of that review.

11:26

An editorial review is important because SPS are often not completely expert at the fighter points of writing.

11:33
An editorial review will produce a clearer, more polished, more professional final product.

11:39
This review is often done by an experienced exam developer.

11:42
Otherwise, it would be done by a professional editor and this may be an opportunity to enforce any house rules such as applying standardized naming, terminology, measurements formats, or using culturally sensitive, concise or professional language.

11:55
Across all items, it is critical that the subject matter expert be involved.

12:00
Either the editor have some knowledge of the area or that a SME reviews the edits because it's possible to damage the items at this step.

12:09
The final review is called a sensitivity review and it should be conducted by a diverse representative panel of subject matter experts, and this review has several similar objectives.

12:20
Promoting diversity by ensuring the exam items reflect a range of perspectives and experiences and avoiding stereotypes and promote representation of diverse groups.

12:29
Ensuring equity and a fair testing experience for all candidates, regardless of their background or identity.

12:34
Fostering inclusion by eliminating any potentially exclusionary or offensive elements, identifying bias by scrutinizing items for unintended biases that could disadvantage any group.

12:46
And finally, ensuring that items are respectful of cultural differences, ensuring that items are neutral and impartial.

12:53
So clearly there can be a lot of SMEs involved and managing that team is very important.

12:57
And this could be a topic of its own advanced course, but I want to give you some essentials.

13:02
First, establish clear, consistent communication channels, probably plural, like e-mail, Slack, etcetera, and ensure that SMEs are aligned about project goals and timelines.

13:12
Second, gives SMEs specific goals, like writing five items from each, and on the Certiverse platform it takes about 3 reviews for each accepted item.

13:22
So I tell authors to cycle between writing one item and then doing 3 peer reviews because it's helpful for item authors to experience both sides of the review process and this cadence ensures that authors get timely feedback.

13:35
You should schedule regular check in meetings to address progress, discuss challenges, answer questions, and provide updates.

13:42
If you are not able to define the house rules for your exam a priori, then they often evolve organically through these meetings when you discuss issues that arise in the other writing.

13:52
And finally, ensure that SME's have access to all necessary resources, including guidelines, templates, and research materials.

14:00
Finally, I want to say a few words about security, item banking, and documentation.

14:05
The security of the exam items should be carefully considered throughout the item writing process, if possible.

14:11
For example, item assignments and reviews shouldn't allow any SME to see all the items in the pool, although that's not going to be possible if you only have a couple SME's.

14:20
An item banking system is a secure, centralized system for storing, maintaining, and managing exam items, and such a system can be enormously helpful.

14:28
The alternative for a lot of programs to use something like a spreadsheet, which can cause all kinds of problems and proper documentation includes item metadata like the authorship information, reviewers, revision history, etcetera.

14:43
Your exam systems should also track item usage and item and form performance.

14:48
A good exam system saves time and effort and enhances exam security, reduces the burden of maintenance, and facilitates continuous improvement.

14:59
So I hope you've enjoyed this video on item writing, which is a critical step in developing fair, reliable, and valid exams.

15:05
Hopefully, you now have a better idea how you would lead and support a team of item authors.

15:10
In the next video, we'll address the next step, which is beta testing your items.

15:14
That's it for this video.

15:15
Thanks for watching.

15:16
We'll see you in the next.