

## Qualifying Exam Score Reports

Webinar held on September 21, 2021 at 6 p.m. CT

Watch the full webinar [here](#).

View the future webinar schedule [here](#).

View previous webinars [here](#).

**PS – Dr. Pooja Shivraj**

**GW – Dr. George Wendel**

**WL– Dr. Wilma Larsen**

**CA – Cariel Apodaca**

**CA:** Okay, so I'm Cariel Apodaca. I'm the Exam Manager here at the American Board of Obstetrics Gynecology. On the ABOG side of things on tonight are Dr. George Wendel, the Executive Director, Dr. Wilma Larsen, our Associate Executive Director in charge of exams, Dr. Pooja Shivraj, our Manager of Research and Psychometrics, and Amanda Novak, Manager of Certification Standards. I can't even get it out tonight. Dr. Wendel, would you like to welcome everybody to the webinar tonight?

**GW:** Sure. I just want to welcome everybody. This is an important webinar tonight. We appreciate you taking the time out of your busy days to learn this new information, and with that, I'll turn things back to you, well I mean Cariel, sorry.

**CA:** So tonight's discussion will be led by Dr. Shivraj. So Pooja, I'll turn it over to you.

**PS:** Perfect, thank you so much.

Thanks.

Okay, so welcome everyone. Thank you all for being here today. As Dr. Wendel said, this is a very important topic to ABOG, so I'm very glad that all of you could be here today. We are going to be discussing candidate score reports, and for those of you who are candidates with us that took our Qualifying Exam, I'm hoping all of you got your score report last week.

So before we get started on what the actual score reports contain, I wanted to start on standards for score reporting. So I've taken this quote out of the standards for educational and psychological testing published by some of the organizations that collaborate together called AERA, APA, and NMCE that put out psychometric standards for score reporting. And what this states is when test score information is released, those responsible for testing programs should provide interpretations appropriate to the audience. The interpretations should describe in simple language what the test covers, what scores represent, the precision/reliability of the scores, and how scores are intended to be used.

And so as we designed our score reports for candidates, what we did was we looked at the psychometric standards published by these organizations, and we looked at each of these and intended to cover each of these within our score reports.

I want to reflect back on what we did provide to candidates and what we are currently reporting. And so on the left is what we used to provide our prior reporting. We used to provide what is called criterion reference information of pass/fails, so based on a criterion of a standard of what the passing score was, we used to provide a pass or a fail to candidates. We also did provide categorical level feedback, so based on the domains, if you were taking the Specialty Qualifying Exam, we provided information on each of the obstetric domains and what percent you scored based off of that. And we did provide them on your portal within the tables as tabular information.

Currently within our score reports, if you think about the standards and talking about that description of how we are scoring and things, we are currently addressing those standards. So we provide a description of how we are scoring in the model we are using. We are doing a candidate-level feedback for you for helping candidates identify their strengths, their weaknesses, as well as an aggregate level in order for candidates to compare themselves to their peers, as well as where that passing score is so for candidates to be able to tell where they are in reference to the passing score.

The idea is for candidates to self-evaluate and for skill improvement, which generally is the purpose of score reporting. And based on the literature of score reports, there is this necessity to integrate data displays as tabular and graphical information, the relevant information that we provide. And so we're trying to hit all of those things.

In terms of how we designed the score reports themselves, it's not something we came up with internally before publishing it. So I want to walk through the process a little bit of how we came up with these reports.

So thinking about the process, we designed it internally, the psychometric team designed it internally, but then it went through an internal review process of our board, and then we conducted focus groups with residents, with fellows, talking about what they liked, what they didn't like about the reports before finalizing them. And within each of those phases, we applied feedback for those reports taking into account each of the comments made by stakeholders in order to go through this iterative design process of our reporting, but also taking into consideration what the wants and the needs were because we knew we couldn't take into account every single want from residents, from fellows, as well as all the other stakeholders we held focus groups with because not everything would yield valid interpretation. So we did take that into consideration while designing score reports.

So now actually getting into what the score reports cover. The idea of taking these screenshots is not for you to read everything that's in the score reports. If you do have the score reports and are a candidate, you've probably read through all of this, but to walk through what the score reports do provide.

So the first page, this is the second page, but the first page provides just a you've passed or you've failed. The second page does provide the description of the scale score going from 100 to 500, what the passing score is. So this is a screenshot from the Specialty exam. The passing score is at 339, so a person that scored below 339 fails. A person that scored higher than that passes.

But it essentially talks about how we standard set the exam, a description of how we scored the exam, as well as the distribution of everyone else's score. So you can see where you lie in comparison to everyone else that has taken the exam in the same year. The good thing, the truly solid thing about scale scores are that if you've taken it the previous year, you can actually compare yourself to your prior year score. So if you've scored hypothetically a 325 the prior

year and a 335 this year, you can actually tell the improvement you've made from the prior year to this year.

Prior to last year, if you'd made a 70 on the last year's exam and then 80% on this year's exam, you can't really tell improvement from last year's exam to this year's exam because the difficulty of the exam could have changed. So the reason for using scale scores is the idea of comparison.

The next page provides category scores, and the category scores are based on the blueprint domains that we use to design the exam. So each of the categories within this table corresponds to how we designed the exam or the blueprint of the exam. So that describes, the little paragraph there describes that, and the table there provides your percent correct on each of these categories, as well as the mean percent correct across all examinees or candidates that have taken the exam.

At the bottom of that page or the table, you also get a link to the bulletin to give you an explanation of what exactly basic sciences comprises of. So within that category, you get a more detailed explanation within the blueprint of what task statements that comprised. So in case you need to take the exam again or for future learning, you would be able to tell what exactly basic sciences means, and so you'd be able to go back into the bulletin to look at that. It also tells you what percent of the blueprint is actually assessed by basic sciences. So that's a little bit more about your percent cap by category within that table.

And finally, this is only specific to the Specialty Exam, and this is only given for Specialty Exams because of the percent of items within each category. And what this graph tells you is your scale scores by domain. And so these domains are broken down just by large categories of obstetrics, gynecology, office practice, and cross content. And this again is specific only to the Specialty Exam.

Where the dark blue line in the middle gives you your scaled score by domain, the box around it, the little light blue box around it, gives you one standard error around your domain score, and then the black diamond tells you what the domain score, the mean domain score for other candidates are. So it sort of gives you this reliability of your score, as well as where other people are scoring. So if I were to take the exam again, I would notice that my obstetrics and gynecology sections, my domains, are weaker than my cross-content and my office practice in

comparison to everyone else. So it's kind of a quick snapshot of where I lie in comparison to everyone else, as well as my reliability around my score.

And so those are the pages we lay out within the score reporting, and as you can see, it's different from what we used to provide before, as well as an explanation of everything else we provide. And so that's just a walkthrough of the score reporting we provide right now. If you have any questions, please email me at [psychometrics@abog.org](mailto:psychometrics@abog.org), but I'm willing to take any questions right now.

The one thing I didn't mention is the scale scores are very similar to what candidates are used to on the CREOG exam. You've probably seen the same use of the scaled scores on your CREOG exam as well.

***Would these domains be available to Program Directors?***

**PS:** Yes. That is a great question. So we also do reporting for Program Directors where we aggregate scores up from candidate reports to Program Director reports. So the domains are available to Program Directors as domains themselves where they are accessible to Program Directors, but Program Directors also get feedback similar to candidates who get feedback on their programs and how they perform similarly to how candidates get reports.

***Will Program Director reports be available by the end of October?***

**PS:** Yes. Program Director reports, the program performance reports, should actually be available within the next two weeks at the latest.

And programs do get aggregate scores and candidate scores, if candidates choose to release their scores to Program Directors.

***Would a Program Director get domain level scores or just overall scores of an individual graduate?***

**PS:** They get overall scores of candidates, not their individual breakdown of scores. So not their domain level scores, they get program level domain scores. So the candidate reports aggregated up to the domain level.

***For a small program, a fellowship that might only have one candidate taking it, would the program report in essence be the individuals performance so that a program would somewhat know how that graduate performed if they were the only one that likely took the test?***

**PS:** Yes.

**CA:** I'm not seeing any more questions. While we're waiting in case anybody has any to add, we will note that this has been recorded. We will distribute it in a couple of days. It usually takes our communications team just a little while to get everything processed because we do try to include transcripts where possible. So as soon as that's ready, we will distribute it to everybody who received the invite for this, and it will, as with all of our other webinars, it's available on the ABOG website.

**PS:** Okay, and if you have any additional questions, please feel free to email me at [psychometrics@abog.org](mailto:psychometrics@abog.org), and I'll be willing to answer them.

**CA:** Alright, thanks everybody. Have a good night.