Past Event: 2021 NCSBN REx-PN Conference **- Standard Setting Transcript**
©2021 National Council of State Boards of Nursing, Inc.

Event
Past Event: 2021 NCSBN REx-PN Conference

More info: https://www.ncsbn.org/16034.htm

Presenter
Hong Qian, PhD, Manager, Measurement and Testing, Examinations, NCSBN

Hello. I'm Dr. Hong. I will talk about standard-setting. How are the passing standard for REx-PN set? After this session, you will know why conducting standard-setting, how important it is, and how the passing standard for the REx-PN was set.

But first, what is standard-setting? It is a process of determining one or more cut scores on a test. If you are familiar with educational tests, they sometime classify students into three levels, basic, proficient, and advanced.

In this case, you need two cuts scores to classify students into three levels. But for licensure exam, we only need one cut to classify students into pass or fail. In July 2021, the registrar and the CEO of the British Columbia College of Nurses and Midwives, BCCNM, and the executive director and CEO of the College of Nurses of Ontario, CNO, approved the passing standard for the REx-PN examination.

The passing standard is -0.32 logits. The passing standard will take effect January 4th, 2022 and will remain in effect through December 2026. Currently, NCLEX-RN passing standard is 0.00 logits and NCLEX-PN passing standard is -0.18 logits.

Does it mean REx-PN standard is lower than NCLEX? Not necessary. REx-PN and NCLEX are different exams, and they are not on the same skill. You may wonder, "The passing standard is -0.32 logits. What does it mean?"

In fixed-form exams, students take same questions. If you get, say, 75% of questions correct, you pass the exam. The passing standard in the percentage correct. But REx-PN uses computerized adaptive testing, CAT, which means students take different questions based on their abilities.

For low-ability students, you work at easy questions. For high-ability students, you will get harder questions. Therefore, percentage cannot be used as a passing standard in CAT exams. Instead, CAT uses current computer technology and measurement of theory to place students on an ability continuum.

Passing standard is a cut point along the ability continuum that marks the minimum ability level requirement. If your ability is above the passing standard, you pass. If your ability is below or at the passing standard, you fail. I will use some terminologies in my presentation.

I will say cut score, cut point, passing standard, or pass mark. They all mean the same thing. For people who set standards, I will call them standard-setting panelists, standard-setting judge, or standard-setting subject matter expert. Why is standard-setting important?

If the standard is set too low, what will happen? Some candidates who are not qualified will be licensed. Do you want to be taken care of by those candidates? Probably not. As an educator, I understand that sometimes you wish the passing standard is lower so all your students can pass.

But we must keep public safety in our mind. What if we set the standard too high? High standard is good. If the standard is set too high, some candidates who are qualified will be denied a license, and we would not have enough nurses. We can't set the standard too low.

We can't set the standard too high. Our goal is to set an appropriate standard which accurately reflects the amount of nursing ability required to practice competently at entry level. Then, how are the passing standard for REx-PN set? Where does REx-PN cut score, -0.32, come from?

Here are the steps for setting the passing standard. One, select the panel of judges. We need people to do it. Two, define minimally competent candidate, which is a key concept when you set a passing standard. Three, train judges on the standard-setting method.

It's not like you just think about a number, there is a systematic way to do it. We need to train them so they understand what they are doing. Fourth, conduct a standard-setting workshop. This is a two-day workshop to set a passing standard. The final step is to establish the passing standard.

I will describe these steps one by one. First step, select the panel of judges. You know it is high-stake decision. Can you guess how many judges are needed? The answer is nine. Minimum of 9 judges are needed, but we had 12 in June 2021 standard-setting workshop.

Panel includes at least one faculty member. Panel includes at least one entry-level nurse with less than 12 months of clinical experience. Another important question is who are qualified as judges? We follow the criteria approved by Practical Nurse Exam Committee, PNEC.

Step two is to define minimally competent candidate. Think about your 20 nursing graduates here from left to right, order them from lowest ability to the highest ability.

Then, I ask you to draw a dividing line from where you're willing to hire them as an entry-level nurse. If you draw a line here, on the left side of the line, their ability is not so much. You wish they had never enrolled in the nursing program at the first place. For the people on the right side, they are good enough.

The most-right one is your top one student. But our focus here is not the ones on the left, not the ones on the far right. Our focus is only one person. Who is the luckiest person that you are willing to hire that just have enough knowledge? The one on the borderline, right?

This is why you'll think a lot, and then you'll say, "Okay, let's draw the line here. I think I can bear with you, train you, and you are safe enough to be an entry-level nurse." This one on the fence is the

minimally competent candidate. To establish a passing standard, we focus on the candidates that possess just enough knowledge to practice entry-level nursing safely and effectively.

We don't focus on students whose ability is very high or very low. We only concentrate on those who are on the fence. In the workshop, we would spend five hours to define minimally competent candidate. Our nurse will go through all the activity statements in eight content areas and look at how a minimally competent candidate would do for each activity.

We even give a minimally competent candidate a name. Who is that student on the fence? Oh, that's Patty. How would Patty do in these activities statements? We spend a lot of time to discuss this concept, so every judge will be on the same page. After minimally competent candidate concept has been built, the third step is to train judges on standard-setting method.

The method is called Modified Angoff Method. In this method, we ask judges, "Imagine, out of 100 minimally competent candidates, how many would answer the item correctly?" Instead of 1 Patty, now, I give you 100 Pattys. They are all minimally competent candidates standing right here and answering the same question.

How many would answer the item correctly? The ratings will be between 1 to 99. If you think this item is so difficult, 100 Pattys, only 1 can answer it correctly, then the rating is 1. If you think this item is so easy, 99 Pattys can answer it correctly, then the rating is 99.

But these are two extremes. Think about that. If you have 4-option multiple-choice item, even by guessing, 25 out of 100 Pattys would get it correct, right? So, most time, the rating will not below 25. On the other hand, the rating will not be, like, 90, 95, because for minimally competent candidates, they're not that competent, right?

If you think an item is really easy, probably, 70 or 80 Pattys would get it correct. If you think item is middle-range difficulty, probably 50, 60. For difficult one, maybe 30, 40 out of 100. Therefore, lower ratings means more difficult item.

Another thing you need to pay attention is would versus should. If you think about should, everybody should answer this item correctly. Too high expectations. Let's face reality. Give you 100 Pattys, how many of them would answer this item correctly?

No should. Use would. Now, treat yourself as a judge. I will show you an item, and you give me a rating. Think about 100 minimally competent candidates, Patty or somebody else. In your mind, you know somebody right on the fence.

Give you 100 of that person, how many would answer this item correctly? Read the item, and give me a rating. If you ask me which one is the key, I will not tell you. If I tell you the key, you will think this item is easy.

And we do the same thing in the real workshop. We do not tell judges the key in the first round. Okay, what is your rating? Sixty? Seventy? Thank you. Now, I tell you, the key is A.

That's what we do in our standard-setting workshop. It's a two-day workshop. Judges will review and rate 120 operational items across all test plan area. You just did one. And our judges will do 120, one by one. These items across all difficulty level and is a representative mini pool.

After they give us a rating, we will ask them to discuss, why you think this item is easy for Pat or why do you think this item is difficult to Pat? After discussion, they would do second-round rating. And we will use second-round rating to calculate cut score.

But you may wonder how to connect these ratings to -0.32 logits. What you get is 60 out of 100, 70 out of 100. It has nothing to do with -0.32 logits. How do you transform these ratings to logit scale?

In REx-PN, it's not like how many percentages you get right, then you pass, because everyone gets different questions. You cannot use a percentage as a passing standard in REx-PN exam. Let me show you how to calculate a cut point. Here, instead of nine judges, I only listed two judges, Judge 1 and Judge 2.

Instead of 120 items, I listed 3 items. But you'll get the idea. For Item 1, Judge 1 thinks 50 out of 100 minimally competent candidates would get this item correct. So, the rating is 0.5.

But Judge 2 thinks it is very difficult item, only 33 out of 100 Pattys will get it correct. So, the rating is 0.33. In our workshop, we will have nine judges and nine ratings for Item 1. For Item 2, Judge 1 thinks it is difficult, and only 25 out of 100 Pattys would get it correct. Judge 2 thinks it is similar to Item 1, and the rating is 0.33.

And so on for Item 3. If you add 3 items together for Judge 1, you get 1.75. For Judge 2, it is 1.16. Then you take average. Add the two numbers together and divide by two. You have cut score 1.46.

For the 3-item test, based on 2 judges, if you get 1.46 out of 3, you pass. It makes sense. For three items, if you get half correct, you pass. But still, how does it transform to logit scale? I know, as a nurse, you are very familiar with the growth chart.

Horizontal axis is age. Vertical axis is height. If I tell you the age, you can tell me the height. On the other hand, if I say the height, you can tell me the age. Now, let us look at unit of measurement. What is the unit of measurement for height?

Inches or centimeters. What is the unit of measurement for age? Years. Now, you have the idea. These are two different things, but you can draw the line and transform to each other. It's similar in the psychometric field. Now, horizontal axis is ability.

Vertical axis is the probability you get item correct. In previous growth chart, you have a line that shows for girls from 2 to 20. The older you are, the taller you are. Here, the line shows if your ability is higher, you have more chance to get the item correctly. Makes sense, right?

If your ability is zero, what is the probability you'll get the item correct? Find the zero on horizontal axis, go up to reach the line, and then go left. It's about 0.4. If your probability is 0.6, what is your ability? About 1.

What is the unit of measurement for ability? Unit of measurement is logit. Now, let's look at this. Horizontal axis is ability, the unit is logit. Vertical axis that goes from zero to three for our three-item test. The minimum you can get is zero, the maximum you can get is three.

The line shows the high ability you are, more likely, you can get three. Do you remember our cut score for the three-item test based on two judges? 1.46. Then, what is the largest skill for this cut? You'll find 1.46 on the vertical axis, go right, reach the line, and then going down.

It's about -1 logit. For REx-PN, we have ratings from 120 items, so the vertical axis goes from 0 to 120. Base on the ratings from 9 judges or more, we draw the line down, it is -0.32 logits for REx-PN.

This is the passing standard recommended by the panel of judge. Then, the final step is to establish the passing standard. PNEC recommended the passing standard to the registrar and the CEO of BCCNM and the executive director and CEO of CNO considering two things.

Of course, the first thing considered is the passing standard recommended by the standard-setting panel. PNEC also considered candidate performance from field testing in 2020 and 2021. PNEC understands if you recommended this standard, how many would pass based on field testing data. We want PNEC to understand the impact of their recommendation.

Then, the registrar and the CEO of BCCNM and the executive director and the CEO of CNO approved the passing standard of -0.32 logits. Take-home message.

Standard-setting is a policy decision, and no true value exists. As a psychometrician, I like true value. Like, 1 plus 1 must be 2. Unfortunately, in standard-setting, there is no true value. Think about that. Are there true differences between -0.32 logits to -0.3199 logits? Not really.

Even though we do standard-setting in very rigorous way, there is no true value. Passing standard is related to mission of the organization and exam. For REx-PN, the mission is public protection. When we set the standard, we need to keep the mission in our mind. Standards should be revisited when test specifications, scope of practice, education programs, and/or other factors change.

When we change a test plan, we need to reevaluate the standards. Thank you very much. I hope the presentation is helpful to you. We will have a live Q&A coming up shortly. You can enter any questions you might have in the Q&A box, and we will answer them later.