Past Event: 2021 NCSBN REx-PN Conference - Computerized Adaptive Testing (CAT) Video Transcript

Event
Past Event: 2021 NCSBN REx-PN Conference

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

Presenter
Shu-chuan Kao, PhD, Senior Manager, Measurement and Testing, Examinations, NCSBN

Hi, everyone. Welcome to the REx-PN Conference, and thank you for joining this session. My name is Shu-chuan Koa. I'm the senior manager of measurement and testing at NCSBN. Today, we are going to cover the Computerized Adaptive Testing, also known as CAT.

In today's presentation, you will understand the concept of Computerized Adaptive Testing and its three important components. What is adaptive testing? Adaptive testing is originally called tailored testing.

It is designed to adjust the level of difficulty to match your ability as a test taker. Every exam is customized for the person taking it. How does it work? The computer algorithms select items from a large item bank for you to answer. The computer aims to identify your ability and give you items that are challenging to you.

If you are a high-ability candidate, you will receive more difficult items because these items can identify how high your ability is. If your ability is low, you will receive more easy items that you can possibly answer.

The purpose of the CAT exam is not to tell the percentage of items you can answer correctly. The REx-PN CAT exam is a different type of exam that estimates your ability in real-time and compares your ability to the passing criteria. Every time when you answer an item, the computer estimates your ability based on all questions you have answered, and it then select an item that is comparable to your ability estimate.

If you answer an item correctly, the computer will find a more difficult item for you to answer. Will the computer just randomly select any items that is more difficult than the item you just responded to? Nope. We program the computer in a way that the next item selected for you will give you a 50% chance to answer it correctly.

When you see this item, you may feel that you can do it, but you are not 100% sure your answer will be correct. Based on the statistical models, the 50% probability of answering an item correctly can give us the most information about your ability. Don't worry, I'm not going to bore you with statistics today.

Instead, I want to emphasize to you that we have a large enough item bank to support the CAT exam. If your ability is very high, we have very difficult items for you. If your ability is very low, we still can find very easy items for you. The CAT exams are challenging for all test takers.

If you take a CAT exam and start to see difficult items, that means your ability estimate is getting high, and you may have a good chance to pass the exam. Every time you respond to an item, the computer re-estimates your ability based on all questions you have answered.

And according to the updated ability estimate, the computer finds you the matching item for you to answer. The more items you answer, the more the computer knows about your ability. Therefore, the computer is able to give you a more precise ability estimate. This is a graphical representation of the adaptive algorithms.

The y-axis is the scale used for both item difficulty and the candidate ability. On this scale, a larger number always indicates a harder question and higher ability. A smaller number always indicates an easier question and a lower ability. The x-axis is simply a tracker of the number of operational questions.

The horizontal blue line is the passing standard. This is the criteria that a computer uses to determine if a candidate should pass or fail the exam. It reflects the ability level of a minimally competent candidate on the REx-PN exam. If you are interested in knowing how the passing standard is determined, please attend the presentation for standard setting.

To begin an exam, the CAT algorithms will select an item that is easier than the passing standard. The blue dot represents the item difficulty of the item selected. When the first question is answered, the computer starts to identify the candidate's ability, which is represented by the black square.

The line above and below the black square is the 95% confidence interval. This is the mathematical range within which the computer thinks the candidate's true ability is. Since the first question is answered correctly, the computer starts to estimate the candidate's ability and then selects a more difficult item to administer, and that the item difficulty of the second item is very close to the updated ability estimate.

After the candidate answers the second question, the computer evaluates how well the candidate has performed on the first two questions. Since the first two questions are answered correctly, the computer gives a higher ability estimate and then selects a more difficult item to administer. Again, this item has the item difficulty very close to the updated ability estimate.

When an item is inserted incorrectly, the ability estimate will get lower than the previous one and an easier item that can match the updated ability estimate is selected. This adaptive process will go on as the exam continues.

Now, I would like to introduce you to the three components of the REx-PN CAT. In the REx-PN exam, all candidates have four hours to answer questions. The length of the exam will vary depending on the

candidate's ability. The minimum nurse exam will have 90 items, which comprise of 60 operational items and the 30 pretest items.

The operational items have the known item difficulties obtained from the pretest, so the CAT algorithms can use these items to estimate your ability. The pretest items are new items. We'll use your responses and ability estimate to determine item difficulties of the new items. The maximum length exam will have 120 operational items and the 30 pretest items.

The exam can stop with minimum number of questions if your ability is far above or below the passing standard. This means when you respond to 60 operational items, the computer knows that you will definitely pass or fail the exam and that there is no need for you to respond to more questions.

If your ability is very close to the passing standard, you will be given a longer exam because the computer needs to collect more information about your ability to make the pass/fail decision. On this slide, I would like to introduce you to content balancing. Content balancing means that all candidate receives the same distribution of items by content area.

For the REx-PN exam, we use Test Plan as a blueprint of all exams. The Test Plan provides the summary of the content and the scope of the licensure examination. No matter how many items you have in your exam, your exam meets the specified percentages of the Test Plan.

Oh, you may wonder how we can assure the content coverage is the same for all CAT exams. We programmed the computer to do two things. Firstly, before selecting the next question, the computer will determine the percentage of items you currently have in all A content areas and to find out which content area deviates the most from the Test Plan.

Secondly, the computer will select an item that targets to your ability from that content area. This way, no matter how many items you have in your exam, your exam can comply with the Test Plan. Starting from this slide, I'm going to show you how the computer decides whether you should pass the REx-PN exam.

In the REx-PN exam, we only report the pass/fail classification. You have four hours to take the exam, and you have to respond to at least 60 operational items. Once you have answered 60 operational items, your ability estimate is compared to the passing standard.

The computer has three different rules to terminate the exam and makes the pass/fail decision. The first stopping rule is the 95% confidence interval rule. After you have answered 60 operational items, the computer starts to check if your ability is clearly above or below the passing standard.

The exam continues when the confidence interval straddles the passing standard. The exam stops when the confidence interval is clearly above or below the passing standard. When the confidence interval of your ability is above the passing standard, you pass the exam, congratulations. When the confidence interval is below the passing standard, then the person failed the exam.

This is a graphical representation of a minimum length exam. Again, the blue horizontal line represents the passing standard. The test stops at item 60 because it is the minimum number of operational questions required. And at that point, the final ability and the confidence interval are below the passing standard.

In this case, the computer concludes that the candidate failed the exam. This slide will show you a different scenario. When you answer 60 operational items, the computer cannot decide whether or not you should pass the exam and continues to select items for you to answer.

The exam stops at item 105, because after this item, your ability estimate and confidence interval are above the passing standard. Yeah, congratulations, you pass the REx-PN exam. Now let's look at a second rule to terminate a CAT exam.

If your ability is very close to the passing standard, this stopping rule may be used to terminate your exam. When you answer 120 operational items, the computer disregards the 95% certainty requirement and then makes the pass/fail decision based on your final ability estimate, which is calculated based on all 120 questions you have answered.

If your final ability is above the passing standard, you pass the exam. If your final ability is at or below the passing standard, sorry, you fail the exam. This example will illustrate a maximum length exam. The computer concludes that you pass the exam because the final ability is above the passing standard, even though the confidence interval is not totally above the passing standard.

If you use all four hours exam time then your exam has not been terminated by either the 95% Confidence Interval rule or the maximum length rule. The running out of time rule will be applied to stop your exam. This means when the exam time is up, the computer again disregards the 95% certainty requirement and then uses your final ability to make the pass/fail decision.

Again, your final ability is calculated based on all questions you have answered. If you have answered at least 60 operational items and then your final ability is above the passing standard, you pass the exam. If you have not answered at least 60 operational items, or your final ability is at or below the passing standard, you fail the exam.

This example will demonstrate an exam that ran out of time. When you answer 60 operational questions, the computer starts to compare your ability to the passing standard. When the exam time is up, the exam stops. The computer concludes that you passed the exam because the final ability is above the passing standard. This is another example.

Please note that this candidate's ability is very high. This candidate failed the exam because he or she does not take enough operational items within four hours. This is the last example of today's presentation. When the exam time is up, the exam stops. The computer evaluates the final ability estimate.

This candidate failed the exam because the final ability is below the passing standard. We have covered a lot of information today. If you want to know more about REx-PN CAT, please visit rexpn.com. Please enter any questions you might have in the Q&A box, and then we will answer them in the next Q&A session.

Thank you for your time.