

# Next Generation NCLEX® NEWS

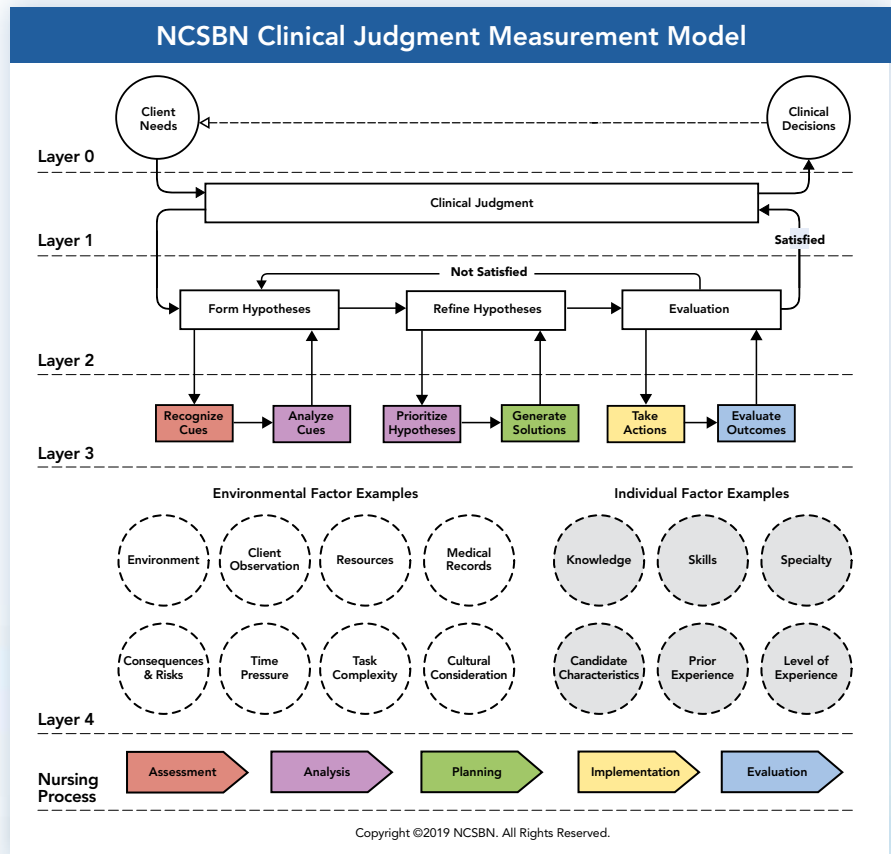
The Next Generation NCLEX® News is a quarterly publication that provides the latest information about the research being done to assess upcoming changes to the NCLEX Examinations. In this issue, you will find information related to the Next Generation NCLEX (NGN) Test Design.

## Next Generation NCLEX®: Test Design

By Joe Betts, PhD, MMIS; William Muntean, PhD; Doyoung Kim, PhD & Shu-chuan Kao, PhD

### Introduction

Previous editions of the Next Generation NCLEX® (NGN) News have provided ongoing information about the state of project development as we approach the launch of the NGN in 2023. This issue provides an overview of the new test design for both the NCLEX-RN and NCLEX-PN exams. The impetus for the NGN is based on the recognition that entry-level nursing practice relies upon a strong foundation of clinical judgment (CJ) and effective decision making to ensure patient safety. This is the motive behind the Next Generation NCLEX — the inclusion of items written specifically to measure CJ through the use of stand-alone items and unfolding case studies. Those unfolding case studies are sometimes referred to as evolving scenarios that follow the **NCSBN Clinical Judgment Measurement Model (NCJMM)**. While the NCLEX item bank has always contained CJ items, those items were not written to specifically target the underlying judgment and decision-making process like the NCJMM.



A number of important frameworks guided the development of the final test design. Most importantly, current professional standards provided the foundation<sup>1,2,3</sup> underlying the entire development process. Secondly, professional guidelines and best practices<sup>4,5,6,7</sup> were followed at every step. Thirdly, decisions on all aspects of the test design were referenced with literature reviews on historical and professional research. This literature review fed into the final framework in the test development process: implementing research through the Special Research Section (SRS) on the NCLEX-RN and NCLEX-PN exams, and through numerous statistical and psychometric simulations. Active research established the psychometric properties needed to validate the design decisions and maintain the exams' high standards of psychometric rigor and validity.

The accumulation of research, along with the aforementioned frameworks, resulted in a final test design. This final test design was reviewed by the NCLEX® Examinations Committee, the Technical Advisory Committee, external professional consultants, the NCSBN Board of Directors and all NCSBN Member nursing regulatory bodies (NRBs). The final test design was approved by the Delegate Assembly representing 58 member NRBs at NCSBN's 43rd annual meeting ([www.ncsbn.org/16129.htm](http://www.ncsbn.org/16129.htm)).

The following sections highlight the differences and similarities between the current NCLEX and the future NGN exam. The NGN exam is designed to capture entry-level CJ understanding more directly. Hence, the NGN in 2023 should be viewed as an extension of the current NCLEX rather than a radically different exam. As the nursing practice changes and evolves, so must the licensure examination that seeks to evaluate minimal competency for entry-level nurses to function safely and effectively. The NGN will extend the current NCLEX to focus on the assessment of CJ as this skill has evolved as a vital aspect of entry-level nursing practice.

## NGN Test Design: Supporting Research & Development

---

Several guiding principles were incorporated into the development of the final test design. As stated previously, the introduction of CJ measurement is the key difference between the NGN and the current NCLEX. To support this new area of measurement, two specific aspects of development were needed: (1) designing an enhanced item development process for writing and reviewing CJ items, and (2) extending the types of items and response structures needed to adequately measure the new domain. The final approach to CJ item development<sup>8</sup> resulted from extensive, iterative experimentation and subject matter expert (SME) feedback from educators, researchers, and practicing clinical nurses who participated in the numerous research panels that spanned over two years.

- 1 American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). Standards for Educational & Psychological Testing. AERA: Washington, DC.
- 2 Association of Test Publishers, & Institute for Credentialing Excellence. (2017). Innovative item types: A white paper & portfolio. ATP: <https://atpu.memberclicks.net/new-publications-from-atp>
- 3 Educational Testing Service. (2015). 2014 ETS Standards for quality and fairness. ETS: <https://www.ets.org/s/about/pdf/standards.pdf>
- 4 Impara, J. C., & Murphy, L. L. (Eds.). (1995). Licensure testing: Purposes, procedures, and practices. Buros Institute for Mental Measurements: Lincoln, NE.
- 5 International Test Commission. (2005). ITC guidelines on computer-based and internet delivered testing. ITC: [www.intestcom.org](http://www.intestcom.org)
- 6 Lane, S., Raymond, M. R., & Haladyna, T. M. (Eds.). (2016). Handbook of test development. Routledge.
- 7 Brennan, R. L. (Ed.). (2006). Educational measurement, 4th ed. Rowman & Littlefield.
- 8 Betts, J., Muntean, W., Kim, D., Jorion, N., & Dickison, P. (2019). Building a method for writing clinical judgment items for entry-level nursing exams. *Journal of Applied Testing Technology*, 20(2), 21-36.

The results of this research supported the development of a method for measuring CJ that went beyond the current approach built around developing single, individual, stand-alone items (see Figure 2 for an overview of the types of stand-alone items that could be seen on the NGN). A new aspect of the NGN exam will be the use of evolving case studies/scenarios to explicitly measure the CJ. These case studies will be comprised of sets of six items, which are presented in sequence according to the steps in the NCJMM. All case studies will begin with a “Recognize Cues” item followed by an “Analyze Cues” item and so on through Layer 3 of the NCJMM until the final and sixth item measuring “Evaluate Outcomes.” The sequential items are constructed so that all information from a previous item in the set is carried through to the subsequent item within the same set. Therefore, all information will be available within each item set and candidates will not need to navigate forwards and backwards between items measuring different aspects of the NCJMM.

Figure 2

### Overview of Item Types on NGN Organized by Response Type Grouping

#### All current NCLEX item types plus:

##### Extended Multiple Response:

- Select All That Apply
- Select N
- Multiple Response Grouping

##### Matrix/Grid:

- Multiple Response
- Multiple Choice

##### Drag-and-Drop:

- Cloze
- Rationale

##### Drop Down

- Cloze
- Rationale
- In Table

##### Highlight:

- In Text
- In Table

##### Bowtie

##### Trend:

- Can contain ALL new item types except Bowtie

When new information is added at one of the steps, this new information will flow through the subsequent items. For example, in actual practice, when a nursing note is updated to reflect the results of a more recent observation, the old note is still available for access. It is unnecessary to memorize the information on one item because in practice, things such as charts and lab information are continuously available. Likewise, this information would be available on subsequent items.



Concurrent with the ongoing development of the item development process, research was undertaken to evaluate a number of item designs and response structures to adequately capture both the fidelity of CJ scenarios presenting in the field and response structures to capture candidates' answers. Usability studies and think-aloud procedures were used to ensure that the final item types to be used on the exam would not introduce any irrelevant variance. Nurse clinicians and educators, along with nursing students, participated in these studies and provided valuable feedback and direction to ensure all item types were simple to use and posed no difficulties when responding. As a result, the final item types had a layout that made it easy to consume information, a formatting that was similar to what is seen in practice, and response structures that were easily manipulated when responding to the content of the items presented. These results indicated that variations in scores were not related to differential understanding or usability of the item layout or response structures, but were related only to the level of understanding and skill displayed by the participants. This provided assurances that the item types would not introduce differences in scores due to the design of the items or the structure of the responses. Instead, they would be based on differences in skill level of the participants.

The expanded range of item and response types allowed for the extended opportunities to measure entry-level nursing knowledge, skills and abilities. An important outcome of this endeavor was to provide a high fidelity set of items that are well-aligned with

nursing practice. Therefore, all the current item types and all the new NGN item types are eligible for administration across the NGN exam (see Figure 2). Candidates can encounter all new item types at any time during their exam, including CJ and all other content areas items. Additionally, all items of an evolving case study can and will use a variety of these different item and response types for different

steps in the NCJMM. For more information on this, please consult previous NGN Newsletters: Fall 2019, Spring 2020, Spring 2021, and Fall 2021 on the [NGN Resources web page](#).

Another important consideration driving the development of the test design was the psychometric and statistical properties. It was necessary to ensure that measuring CJ did not compromise the psychometric properties of the exam—they had to be equivalent or better than the current exam. The NGN test design met or exceeded this requirement for all important statistical and psychometric properties.

In addition to maintaining the psychometric rigor of the exam, it was important to evaluate the extent to which the length of the exam would need to change. A guiding principle was

to meet or exceed the psychometric rigor but not substantially increase the amount of time candidates would need to complete the exam. Therefore, the NGN test design was finalized because it maintained the psychometric and statistical rigor and it did not appreciably increase the testing time. Consistent with the current NCLEX response times, NGN CJ items will take approximately one to two minutes to complete.



*The expanded range of item and response types allowed for the extended opportunities to measure entry-level nursing knowledge, skills and abilities.*

Another change from the current exam will be the methodology for scoring. Currently, items on the exam are scored as either all correct or all incorrect. There are no credits for partial understanding. With the introduction of the NGN, partial understanding will be evaluated by polytomous scoring models that allow for individual items to be scored for partial credits. For example, the current NCLEX multiple response (MR) select-all-that-apply (SATA) items are scored such that the only way to get a correct score is to endorse all the keys. Endorsing a subset of the keys will result in obtaining an incorrect score. This approach does not differentiate individuals that might get one, two or some other number of correct keys. For instance, if there are three keys and person A gets one correct while person B gets two correct, they are both scored as incorrect at present. However, it seems reasonable that person B showed some higher understanding than person A, but not as much as someone getting all keys correct. The partial credit models will allow for this type of differentiation in candidates' knowledge, skills and abilities.

To support the assessment of partial knowledge, polytomous scoring models will be employed. The [Summer 2021 NGN Newsletter](#) provides extensive information about how these will be applied when scoring individual items<sup>9</sup>. These scoring methods are named the plus/minus, zero/one, and rationale methods. They will be used to transform the options selected by candidates into a raw score for the item that can range from zero to the maximum number of keys. However, the NGN will also employ an item response model<sup>10</sup> that will transform those raw scores into scores on the NCLEX scale. Today, the NCLEX uses the Rasch<sup>11</sup> measurement model for this scaling. The NGN will use a model that expands the Rasch model and is, in fact, a generalization of the model called the Partial Credit Model (PCM)<sup>12</sup>. Using the PCM for scaling the NGN will allow for the scores to be transformed to the current NCLEX scale. This means that score pre-launch and post-launch of NGN will be comparable and decisions about pass/fail status will have been assigned based on the same scale.

*With the introduction of the NGN, partial understanding will be evaluated by polytomous scoring models that allow for individual items to be scored for partial credits.*

9 Also see: Betts, J., Muntean, W., Kim, D., & Kao, S-c. (2021). Evaluating different scoring methods for multiple response items providing partial credit. *Educational and Psychological Measurement*, 1-26: DOI: 10.1177/0013164421994636

10 See volumes in the set: van der Linden, W. J. (Ed.). (2016). *Handbook of item response theory*. Chapman and Hall/CRC.

11 Rasch, G. (1980). *Probabilistic models for some intelligence and attainment tests*. Chicago: University of Chicago Press.

12 Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika*, 47, 149-174.

## Operational Test Design: Similarities and Differences from the Current NCLEX

This section will provide an overview of the dynamics of the operational NGN exam. It is intended to provide information about how the actual exam will be structured, delivered and scored. It will also highlight the similarities and differences between the NGN and the current NCLEX. Figure 3 provides a high-level comparison across a number of important factors.

Figure 3

Overview of Test Design Similarities and Differences between the Current NCLEX and the Future NGN		
Design Specification	NCLEX Today	Next Generation NCLEX (NGN)
Time Allowed	5 hours	5 hours
Delivery Method	Variable-length CAT	Variable-length CAT*
Pass/Fail Decision Rules	CI, ROOT, Max Items	CI, ROOT, Max Items
<b>Total Exam Length (min-max)</b>	<b>75-145</b>	<b>85-150</b>
NCJMM Case Studies	N/A	3 (18 items)
Stand-alone Items	60-130**	52-117**
<b>Total Scored Items (min-max)</b>	<b>60-130</b>	<b>70-135</b>
Unscored (Pretest) Items	15	15

\* Items within a Case Study are static, not adaptive

\*\* Range possible due to variable length CAT

The table provides a comparison between the current NCLEX and the NGN in 2023. Candidates will continue to have up to five hours to complete their exams and their response time per item will be similar to the current timing. All current accommodations will continue to be available. Candidates will be able to follow the same process as they do today to request testing accommodations.

Additionally, the exam will continue to be delivered as a variable-length computerized adaptive test (CAT) which will allow for the exam to select items and now item sets, i.e., evolving case studies, that are optimal for each candidate. Like the exam today, each item is selected to target the candidate's ability level. Candidates at all ability levels will get items specifically targeted at their own levels throughout the test. This means that no matter where on the scale candidates fall, they will all see items that are of a similar difficulty for them. For instance, with the current Rasch model, candidates will have roughly a 50% chance of getting the correct answer on each item. Therefore, no matter where a candidate is on the ability scale, above or below the passing standard, an appropriately difficult item will be selected to target their ability. In this way, candidates get items and cases that are appropriate for their current ability estimate.

The change for the NCLEX will be the introduction of the partial credit scoring model for polytomous items. However, the CAT mechanism will remain the same for both stand-alone items and evolving case studies. Stand-alone items will still be selected to be optimal for each candidate at their unique ability level. Likewise, item sets will be selected to be optimal for each candidate. Specifically, across all six items, each candidate should be able to get about half the total points. This is similar to the current 50% probability selection of

stand-alone items, but applied to items in the set with multiple score points. The only thing to consider is that it is possible for one of the items in the set to be either more or less difficult for the individual, but over all six items, the sets will be selected to be optimal for each individual. Like stand-alone items, this allows for item sets to be selected to target each candidate's unique ability level and ensure that no candidate gets sets that are beyond their current ability range.

As the NGN will continue to be at variable length, there will be both a minimum and maximum number of items candidates will be able to take. Like today, all candidates must take at least a minimum-length exam to have a pass/fail decision to be made. If a candidate does not take at least a minimum-length exam, the exam will be scored as a fail, as it is currently. Today, candidates must take 75 items before an exam can begin to be scored. With the NGN, this will increase slightly to 85 items. Additionally, the maximum number of items will slightly increase from 145 to 150.

What this means is that after 85 items, the NGN scoring mechanisms will begin to evaluate whether or not there is enough information to make a pass/fail decision. If there is not, then the exam will continue delivering items until a stopping rule is reached. The NGN will employ the same decision rules as today to make pass/fail decisions. These are the 95% confidence interval rule, the run-out-of-time rule, and the maximum items rule. These will be implemented exactly as they are currently implemented with no changes. More information can be found [here](#).

The composition of the minimum-length test will be similar to how it is structured presently, with a slight adaptation of incorporating CJ items and

evolving case studies. Within the minimum-length exam, there will be 52 stand-alone items and three case studies (each case has six items for a total of 18 items) that will be scored and contribute to the pass/fail decision. This results in 70 scored items. Scored stand-alone items and case studies will continue to be selected by the CAT algorithm to select the most optimal item or set for each candidate given their current ability estimate.

The remaining 15 items will be comprised of pretest items which is the same number in the current exam. These 15 items can come in one of three combinations: (1) all stand-alone items, (2) one

case study (six items) and nine stand-alone items, or (3) two case studies (12 items) and three stand-alone items. This ensures that all candidates will have an equal number of unscored items in their exams. These unscored, pretest items are used to ensure items in the banks are current with respect to content and psychometric values.

Stand-alone items will continue to be balanced according to

the test plan percentages. For example, currently there are eight content areas in the NCLEX exam, with different percentages of items from those areas making up the exam. Management of care comprises between 17% and 23% of the exam. Because CJ is an integrated process, these items will be selected as a fixed probability. Currently, this is set to be about 10% probability of getting a CJ stand-alone item and if it is not selected, the exam will then follow the content area distribution selection, just as it does today. Note that this percentage, like the content blueprint percentages, could change based on results from future practice analyses which are done periodically to ensure the test blueprint aligns with the frequency and importance of entry-level tasks in daily practice.

*The composition of the minimum-length test will be similar to how it is structured presently, with a slight adaptation of incorporating CJ items and evolving case studies.*

Regarding scored case studies, each candidate will see one in the first third, one in the middle third, and one in the final third of the minimum-length exam. In this way, all candidates will see case studies around the same position in the exam. However, these will not be at the exact same place but vary in their position for every candidate. One candidate might get the first case study after the sixth item, while another might get it after the twelfth, and so on. Case studies will only be seen within the minimum-length test, that is, before item 86 is reached.

After the minimum-length test, if a stopping rule is not triggered, then the exam will continue as it does today. At this point, the exam will select only stand-alone items to help target the individual's ability more directly. This method allows for gathering the most information with each item to help make the pass/fail decision. As each individual item can be chosen at the candidate's current ability estimate, maximum information can be obtained from each item. This precision helps target the candidate's ability and makes the most accurate decision relative to the passing standard.

The exam will continue until a stopping rule is found and a pass/fail decision can be made. For a pass/fail decision to be made, candidates could take one to 65 items in addition to the minimum-length exam (85 items) before reaching the maximum number of items (150 items). One important note is that after every item is answered, the CAT algorithm will use all the responses from all the scored items taken up until that point to update the candidate's ability estimate. That updated estimate is what is used for making the pass/fail decision.

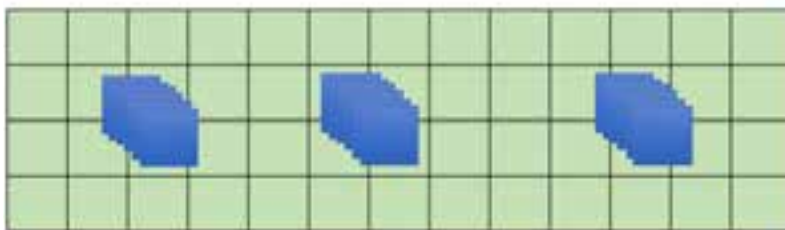
## Example of the NGN Exam

Let's walk through an example to get the feel for what the exam will look like. Let's start off with the minimum-length test. Figure 4 provides an illustration of a minimum-length exam.

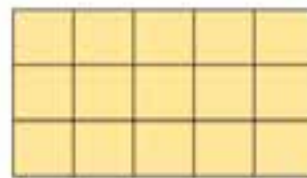
Figure 4

### Example of a Minimum-length Exam

**Scored: 52 + 3 case studies (18 items)**



**Unscored: 15**



**Total: 52 + 18 (3 cases) + 15 = 85**

Here, we have 52 stand-alone scored items represented by the green squares and three embedded case studies of six items apiece (the blue squares overlapping each other). Additionally, there are 15 unscored items, and in this example, each of the yellow squares represents a single, stand-alone item; therefore, this example shows a situation where a candidate would not get any pretest case studies. Taken together, this results in the 85-item minimum-length test. As previously mentioned, any of these items can be any of the

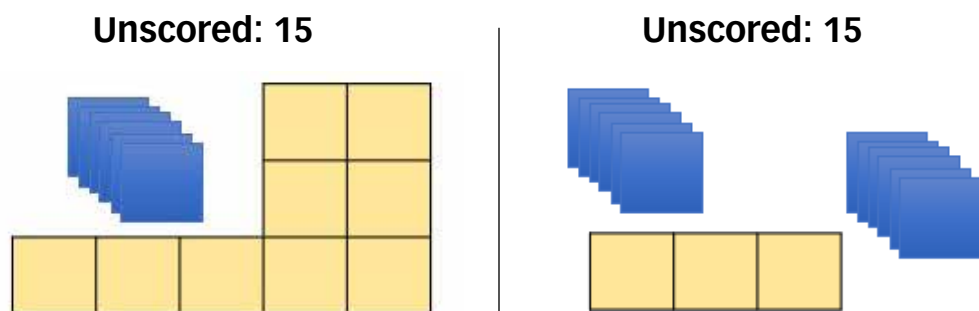


item types highlighted earlier. The case studies will be randomly selected within the first, second and final third of the minimum-length exam. The unscored, pretest items will be randomly selected at any time within the minimum-length exam. All scored items (and item sets) will be selected to be optimal for each candidate. After each item (or item set), the candidate's ability estimate will be updated to reflect the most recent estimate based on all the information from all scored responses and the next item or set will be selected.

According to Figure 4, this looks like a situation where a candidate is selected to have all 15 unscored items as single stand-alone items. However, there are two other possible configurations of the unscored items (Figure 5). The figure shows, on the left, an example where of the 15 unscored items, six of them will be taken up by a case study and the remaining nine will be the stand-alone items. On the right side of the figure is an example where 12 of the items are dedicated to case studies and only three stand-alone items.

Figure 5

### Examples of Different Distributions of Pretest, Unscored Items



Each candidate will randomly get one of these three examples. Therefore, at the completion of the minimum-length exam, all responses up until that point will be used to generate a score for the candidate. The score will be compared to the pass/fail decision rules and if the candidate's score meets the decision criteria, then a pass/fail decision can be made.

However, if a stopping criterion is not met, then the exam will continue. As the exam is a variable-length CAT, after each item the exam will evaluate and update the candidate's ability estimate. When one of the stopping rules is triggered, a pass/fail decision can be made. This means that the candidate could take up to 65 stand-alone items post the minimum-length exam (85 items). Again, this is also why only stand-alone items will be used in this part of the exam because it could terminate after any additional item.

After the minimum-length exam, the NGN will select items in a two-phase approach. First, each item will have a 10% probability of being selected for a CJ item. Thus, out of 65 items, we would expect about six or seven items to be CJ stand-alone items. If the random selection picks a CJ item, an optimal stand-alone CJ item will be selected for the candidate.

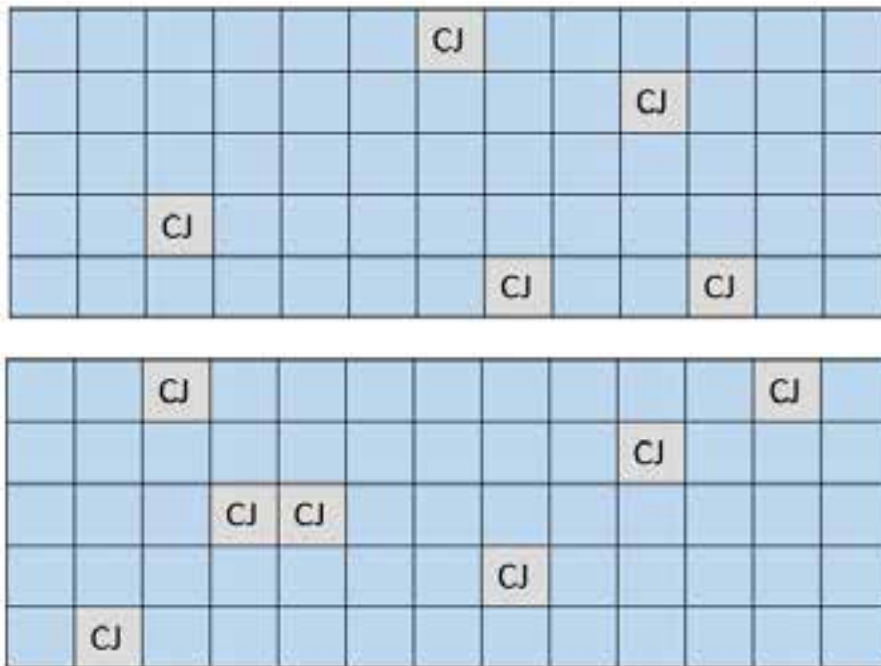
However, if it does not indicate a CJ item should be selected, then the selection algorithm will move to the second phase. Now it will select items based on the test plan category percentages from the content area blueprint. This is itself a two-step method. It will begin by evaluating the number of items given in each

content area and compare that to the test blueprint. The content area that is the furthest out-of-balance with the blueprint expectations will be chosen for item selection. Then an item from that content area will be selected from a set of optimal items targeting the current ability estimate of the candidate. This two-phase process will continue until a stopping rule is triggered.

Figure 6 provides some examples of hypothetical maximum-length exams. Here, the squares represent the 65 stand-alone items after the minimum-length exam. The blue squares represent the content area items and the grey with CJ items. The top example shows an exam with only five CJ items selected and the bottom example shows one with seven.

Figure 6

### Examples of Maximum-length Exams



## NGN Score Reporting

Each candidate will continue to be provided with a Candidate Performance Report (CPR). If a candidate passes the exam, the report will look almost exactly as it is today. If a candidate takes at least a minimum-length exam but fails the exam, the report will look similar to the current CPR; however, it will be expanded slightly. The report will continue to provide diagnostic categories of how the candidate performed on each content area related to the passing standard and whether they are above, below or near the standard on each content area. The main change will be an addition of the CJ performance. The additional CJ diagnostic reporting will be comprised of a single overall CJ score that indicates the candidate's performance on all CJ items. In addition, score categories will be reported for each element in Layer 3 of the NCJMM.

## Summary

The NGN will remain very similar to the current NCLEX. There will be a similar number of items using the variable CAT methodology. The same stopping rules will be applied to determine pass/fail status; these stopping rules follow the same methods as they currently do. All items (and now sets of items for evolving case studies) will continue to be selected to be optimal for each candidate. This is to ensure that no candidate gets items (or item sets) that are too easy or too hard, regardless of the candidate's ability. Candidates will follow the same process as the current NCLEX to request testing accommodations.

The biggest changes will involve the direct measurement of clinical judgment and the use of evolving case studies measuring Layer 3 of the NCJMM. This will also entail the use of a number of new item and response types that allow for expanding the measurement of all aspects of the exam. This allows for greater breadth of coverage of all entry-level nursing knowledge, skills and abilities. Again, there are several resources available to familiarize oneself with the new item types. There will be an online tutorial, similar to the current one, that will enable candidates to interact with the item types as much as they would like before their exam day. Finally, the new item types will no longer be scored as either all right or all wrong. Instead, the items will be scored polytomously with multiple score points to allow for the evaluation of partial understanding and to better assess each candidate's entry-level nursing knowledge, skills and abilities.

## NGN Resources

For more information regarding the NGN project, please visit the [NCSBN website](#) and our [Frequently Asked Questions](#), which address common questions from candidates and educators. The [NGN Resources](#) page includes past publications of the NGN News. The newsletter is published quarterly and provides the latest information about the work to assess potential changes to the NCLEX Examinations. [NGN Talks & Videos](#) houses short [NGN videos](#) on topics related to the NGN.



Next Generation NCLEX® News is published by  
**National Council of State Boards of Nursing (NCSBN)**



111 E. Wacker Drive, Suite 2900  
Chicago, IL 60601-4277

Phone: 312.525.3600  
International Calls: +1.312.525.3600

Website: [www.ncsbn.org](http://www.ncsbn.org)

*NCSBN provides education, service and research through collaborative leadership to promote evidence-based regulatory excellence for patient safety and public protection.*

Copyright ©2022 NCSBN. All Rights Reserved.