Journal of Applied Testing Technology, *Vol 20(S2), 21-36, 2019*

Building a Method for Writing Clinical Judgment Items for Entry-Level Nursing Exams

Joe Betts[1*], William Muntean[1], Doyoung Kim[2], Natalie Jorion[1] and Philip Dickison[2]

[1]Pearson VUE, Chicago, IL; joe.betts@pearson.com
[2]National Council of State Boards of Nursing, Chicago, IL

Abstract

Clinical judgment has become an increasingly important aspect of modern health service professionals. To ensure public safety, licensure exams must go beyond assessing only knowledge and skills when evaluating entry-level professions to evaluating clinical judgment. This importance necessitates licensure and certification examinations in these professions to evaluate the extent to which this domain can be measured. This study will provide background of a large-scale licensure examination's process for measuring clinical judgment in entry-level nursing. As clinical judgment is seen to be a higher-order, more complex construct than measuring basic knowledge, an outline of the process that was used to develop the construct and build items to a specific task model will be highlighted. In accordance with existing adult learning theories and contemporary nurse program delivery, the process of specifying the requisite cognitive theories guiding nursing practice will be elucidated and interwoven with the item development discussion. The item writing and review process will be described in detail and would be generalizable to professions outside of nursing.

Keywords: Assessment Design, Higher-Order Cognitive Construct, Information-Processing Framework, Nursing Clinical Judgment, Task Model, Technology Enhanced Item

1. Introduction

Over the past decade, the field of nursing has seen a growing expansion of the necessary skills that nurses must perform competently and safely. The most predominant skill needed is clinical judgment, a skill that applies and makes use of basic professional knowledge. Clinical judgment is a higher-order cognitive construct that represents medical knowledge, skill, decision making, and critical thinking (Dickison, Luo, Kim, Woo, Muntean, & Bergstrom, 2016; Muntean, 2012). In the case of nursing, the construct goes beyond the acquisition of nursing knowledge and is defined as the active use of nursing knowledge when making decisions and judgments about many aspects of patient situations (e.g., client/patient need, health concerns, presenting symptoms, etc.; Tanner, 2006). Clinical judgment is a combination of critical thinking (Rubenfeld & Scheffer, 2010) and decision-making skills (Thompson & Dowding, 2009) that is based on a solid foundation of basic professional medical knowledge, in this this case with respect to nursing knowledge (Kuiper, O'Donnell, Pesut, & Turrise, 2017; Standing, 2014) and is critical to the general nursing process (Wilkinson, 2012).

The changing landscape of health care across the United States contributes to the growing emphasis on clinical judgment. Nurses are gaining greater responsibilities and accountability (Casey, Fink, Krugman, & Propst, 2004; Ebright, Urden, Patterson, & Chalko, 2004; Hickey, 2009; Saintsing, Gibson, & Pennington, 2011; Simmons, Lanuza, Fonteyn, Hicks, & Holm, 2003), and they are dealing with increasing levels of patient acuity in their respective organizations. This comes from an aging population and emerging clinical interventions that prolong life and facilitate recovery from serious injury and disability. In all, individuals are living longer and living with greater levels of disability than previously, and entry-level nurses are at the front line when caring for and treating these individuals.

Another factor contributing to the increased importance of accurate clinical judgment is the association

*Author for correspondence

between poor patient outcomes and poor nursing clinical decision-making (Nibbelink & Brewer, 2018). Berkow, Virkstis, Stewert, & Conway (2008) reported that only 20% of employers were satisfied with the decision-making abilities of new nurses. For that reason, more educators are incorporating clinical judgment in the core curricula of entry-level nursing courses (Berman & Snyder, 2012; Craven, Hirnle, & Henshaw, 2017; Potter & Perry, 2009). It provides a foundation for the nursing process that is crucial for competent and safe nursing (Berman, Snyder, & Frandsen, 2016; Burton & Nay Ludwig, 2015; Potter, Perry, Stockert, & Hall, 2017; Taylor, Lilis, Lynn, & Le Mone, 2015; Wilkinson, Treas, Barnett, & Smith, 2016).

It is now expected that the entry-level nurse will not only have the prerequisite knowledge and skills to function competently and safely but also be able to process information at a higher level to make sound decisions about patients (Lasater, 2007). For that reason, greater importance is being placed on the entry-level nurse's ability to organize complex information, think critically about numerous aspects of the patient situation, and then make correct decisions about how to proceed—they must demonstrate good clinical judgment skills. However, despite the training efforts of educators, studies continue showing that there is a lack of adequate clinical judgment among newly graduated nurses (Kavanaugh & Szweda, 2017).

These trends pose educational and psychological measurement problems. From an educational perspective, they raise the question as to whether educational assessments of clinical judgment are useful in determining competency. Because entry-level nurses have considerable challenges handling the nature of clinical practice that is filled with ill-defined situations and multi-layered presentations (Lasater, Nielsen, Stock, & Ostrogorsky, 2015), educational assessments may not sufficiently capture the intricacy of actual practice. Additionally, many of these types of evaluations can be subjective and ad hoc. Therefore, it would be important to investigate the extent to which this construct can be elucidated and evaluated with an eye towards building a sustainable model for on-going evaluation of the necessary aspects of competent clinical judgment. This type of model could potentially be built into a coherent, structured system of on-going learning and assessment.

Likewise, from a psychological measurement perspective, it is imperative that professional licensing and certifying examinations in the medical field begin evaluating how to assess this higher-order cognitive construct of clinical judgment. This is true for most health care professions, and it is especially true for the nursing field. Alongside evaluating whether candidates show enough basic knowledge and skills, credentialing/licensing examinations must also focus on whether candidates have the necessary clinical judgment skills. Entry-level nurses must demonstrate their ability to manage increasingly complex patient scenarios so that they can make correct decisions and practice competently and safely upon entry into the profession. This requires innovative, high-fidelity clinical judgment assessments and possibly new approaches to ensure that the assessments are measuring clinical judgment.

This study discusses the item development process, outcomes, and results of a project focused on measuring clinical judgment in entry-level nursing (nurses with less than one year of practice) within the context of licensure/certification examinations. Figure 1 describes the general flow of the assessment project, beginning with construct development, transitioning into item development (prototype and functional), and ending with scoring models. We augmented this outline with Dickison, Luo, Kim, Woo, Muntean, and Bergstrom (2016) methodologies for implementing an omnibus clinical judgment framework. The overall approach to this research fits well within the context of a principled assessment design framework (Nichols, Korbrin, Lai & Koepfler, 2016). Although the project took on many interrelated objectives, this article details on one aspect related to item development focused on the development and use of the task model. As defined by Dickison, Haerling, and Lasater (2019), the task model seeks to "facilitate the development of highly structured items that elicit responses and generate data from the test takers in a consistent manner".

2. Measuring Clinical Judgment in Nursing

Measuring a construct begins with a defining statement and a guiding definition, i.e., what is the object/construct of measurement. In this case, the construct is clinical judgment within the domain of nursing for entry-level practice. The initial construct definition considered the entirety of the literature review done with respect to clinical judgment, decision making, and critical thinking in nursing. Taking a broad approach allowed for the presen-

tation of many nuances of the clinical judgment construct to panels of Subject Matter Experts (SMEs) and let those expert nurses determine the importance of these aspects. Through various iterations, the following concise guiding definition of the underlying construct of nursing clinical judgment was accepted:

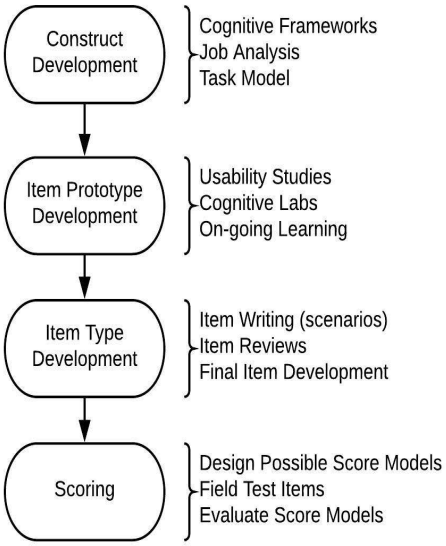

Figure 1. Item development process.

"Nursing clinical judgment is the observed outcome of critical thinking and decision-making. It is an iterative process that uses nursing knowledge to observe and access presenting situations, identify a prioritized client concern, and generate the best possible evidence-based solution in order to deliver safe client care."

This definition succinctly captures the confluence of critical thinking, decision making, and knowledge components to the process. It recognizes an iterative mental process, with the number of iterations representing the number of times a set of judgments are needed when issues evolve over time or new symptoms emerge. Most importantly, judgments are based within the context of providing safe client care using established, evidence-based solutions.

In addition to generating a construct definition, the process by which item development proceeded was guided by two aspects to help represent the construct for measurement (Embretson, 1983). First of all, a strategic job analysis was completed on knowledge, skills, and abilities (KSAs) needed during the initial year of nursing practice (National Council of State Boards of Nursing, 2018). Results from a linkage analysis (Raymond, 2016) within the strategic job analysis identified relationships between all the KSAs evaluated. Nursing process was identified as the most linked knowledge statements and clinical judgment as the most linked skill underlying all of the KSAs (National Council of State Boards of Nursing, 2018). This demonstrates the underlying importance of understanding the nursing process and making skilled decisions for competency in entry-level nursing. These results also further substantiate the previous literature indicating the importance of clinical judgment in the everyday practice of nursing.

The second area of construct representation focused on the main theories of clinical judgment found in the nursing literature to date and will be discussed in the following sections. These theories along with research on decision making in nursing (Muntean, 2012) helped drive the development of the task model framework for item development whereas the strategic job analysis outlined the set of necessary KSAs related to clinical judgment and outlined a set of task statements relevant to practice. In nursing, the understanding of clinical judgment has been based on three fundamental cognitive theories (Muntean, 2012). These are the intuitive/humanistic (Benner, 1982; Tanner, 2006), cognitive continuum (Harbison, 2001) and information processing (Oppenheimer & Kelso, 2015) theories. All three of these theories were used in the construction of the task models used to design items measuring clinical judgment (Dickison, Haerling, and Lasater 2019).

2.1 Use of the Information Processing Theory in Task Model Development

The information processing theory provided a well-validated model for structuring decision making. This model has been used to underpin numerous judgment and decision making situations across multiple professions (e.g., accounting: Libby, 2017; Ismail & Trotman, 1995; marine power plant: Su & Govindaraj, 1986; manufacturing troubleshooting: Schaper, 1998; Bereiter & Miller, 1989). It provides an analytical, hypothesis-driven approach for parsing a problem situation into the requisite cognitive elements needed for decision making. For the current item development endeavor, this is a suitable model because the level of analytical representation fits within the other two leading clinical decision-making theories (see the humanistic and cognitive continuum theories).

The information processing theory is particularly useful for measurement because it reduces the complexity of the cognitive process underlying decision-making by specifying the dominant subcomponents of clinical judgment. It is a structured and comprehensive model that abstracts nursing situations into the requisite elements needed to make competent and safe decisions. Although independent, the components are related to and influenced by one another. Most importantly, each component is potentially assessable. Thus, the source of poor clinical judgment can be potentially traceable to a specific mental process. Identifying the sources of errors in judgment is useful in establishing decision-making competency.

Dickison, Luo, Kim, Woo, Muntean, and Bergstrom (2016) adopted the core principles and diagnostic reasoning approaches described in Elstein, Shulman, Sprafka, and Allal (1978) and developed a hierarchical representation of nursing decision making. This representation offers varying levels of abstraction through latent attributes associated with the use of clinical judgment in nursing (for detailed description, see Dickison, Luo, Kim, Woo, Muntean, & Bergstrom, 2016; Dickison, Haerling, and Lasater 2019). Guided by several research panels and development projects, the current clinical judgment model takes on the structure in Figure 2. Because layer 3 provides a good balance of abstraction and the component processes underlying decision-making, it is suitable for measurement and item development.

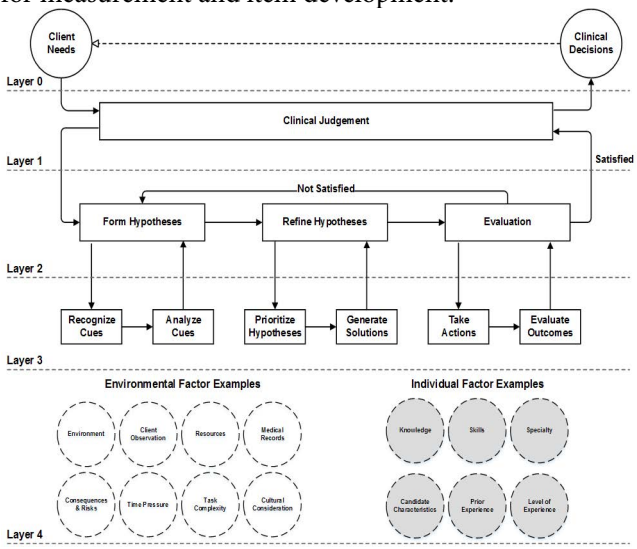

Figure 2. Layered clinical judgment model (Reprinted with Author's Permission).

The clinical judgment process in layer 3 begins with recognizing cues. During the initial presentation of a clinical problem, *cue recognition* describes the mental process involved in extracting and identifying relevant and important information from the presenting situation. This occurs through various forms of observation, whether through the environment, or medical records, symptoms, or vital signs. The key is differentiating between normal and abnormal presenting symptoms/observations/facts. Once these cues are attended to, they are analyzed under the *analyze cues* stage of the model. Cues are analyzed through a number of organizing mechanisms, such as clustering and linking related information to create constellations of individual cues. Hypotheses and holistic appraisals are developed to understand the causes of the clinical problem.

The collection of hypotheses and general appraisals are then prioritized in *the prioritize hypotheses* stage of the model. Given the appraisal of information the process of evaluating and ranking the potential causes or risk factors takes place. Prioritizations takes on many forms and addresses needs such as, what must be dealt with first, what is the most likely cause of the current condition, what is this patient's greatest risk factor, and so forth. The top priorities are addressed by the *generate solutions* process, where one generates a set of feasible solutions to handle the emergent concern. Under conditions of uncertainty when several strong competing hypotheses are indistinguishable given the current state of information, refining the set of hypotheses is necessary. Thus, the generated solutions may also focus on ruling out certain hypotheses (e.g., take further measurements of physical status inquire more deeply into the history of the evolution of the symptoms, etc.). This is part of the iterative process of clinical judgment.

Once a set of solutions is in place, it is time to *take action*. This process involves the implementation of the solutions (e.g., how an intervention should be performed, what communication strategy would be best for this patient, how to best document interventions performed, etc.). Because clinical judgment is a higher-order cognitive skill, the implementations of solutions are multifaceted and are made complicated by confounding circumstances of the problem. For example, consider two patients, Patient 1 and Patient 2, who both have heart failure, but Patient 2 also has 2nd-degree heart block. Both patients require the same solution – management of heart failure – but different implementations. It is appropriate to administer a beta blocker for Patient 1 but not for Patient 2 because it could cause further bradycardia and

potential dysrhythmias. Medical procedures and actions are often taught in isolation, and although they are the building blocks of medical knowledge, they are unlikely to occur in isolation in actual practice and therefore unlikely to represent the true ecology. For that reason, the implementation of the solution under realistic complications is pivotal to clinical judgment.

Finally, *evaluating outcomes* of the actions taken is necessary. Reflecting upon the outcomes of the interventions is important in the decision-making process. Evaluating the correspondence between observed and expected outcomes determines the success of the decision-making process. Thus, understanding what signs point to an improving patient status or declining health is imperative. When the intervention is not effective, the judgment process undergoes iteration, heeding new information to help identify alternative options that could be more effective.

These six core mental processes serve as the basis for item development and the foundation of the task model (the second column of Figure 3). The task model facilitates item development by requiring item writers to specify the expected behaviors of each clinical judgment element. For example, the expected behaviors of recognizing cues are recognizing abnormal versus normal information, recognizing signs and symptoms, or identifying past medical history. An item writer defines these behaviors for the clinical problem they are writing about, and this ensures items adhere to the overarching construct and the operational definition of its respective clinical judgment element.

2.2 Use of the Humanistic Theory in Task Model Development

The intuitive/humanistic theory (Benner 1982) provides a comprehensive set of insights into the critical thinking and decision making of individuals as they develop over time from nascent novices to experienced experts. As individuals acquire broader skills and expertise, they are able to more competently deal with situations as they emerge in a, typically, more fluid manner. The theory proposes five sequential levels of attainment: novice, advanced beginner, competence, proficiency, and expertise.

In regards to decision making, *novices* make clinical judgments that are context-free; they ignore the idiosyncrasies of the situation. This results in decision-making that is primarily rule-based. It is inflexible and results in very limited performance. *Advanced beginners* incrementally account for more situational variables and decision-making attributes gradually become context dependent. Decisions from those in the *competence* stage involve organization structures and are made with greater efficiency, albeit still relying on conscious, abstract, analytical, and deliberate planning. In the remaining two stages, *proficiency* and *expertise*, decisions are based on holistic thinking, with problem features viewed as salient or irrelevant. Situations are organized and analyzed intuitively, and experts act naturally and often reach conclusions without explicit understanding.

The humanistic theory is complementary to using the information processing framework for item development because it postulates that before becoming an expert decision-maker, decisions are reached through analytic methods and systematic planning. In accordance to this theory, entry-level nurses process decisions from the advanced beginner to the competent stages of acquisition, thus lending support for using a process model to measure and assess clinical judgment ability. For safe clinical practice, entry-level nurses must have progressed beyond the novice level that deals with explicit facts and simple rules-based contingencies. Entry-level nurses should be transitioning from the advanced beginner during the first year of practice because along with their clinical experience, they have experience from mentored practice in their educational programs. As a result, their nascent organizational structures bring dynamic aspects of client presentations to bear on the clinical judgment situation.

Within the context of item development, item writers crafted patient scenarios that entry-level nurses typically encounter and that are within the level of their acceptable responsibility. Because the scenarios represent realistic practice, common naturally-occurring complexities were prevalent in the written clinical situations. This prevents using simple repetition of knowledge statements to solve clinical problems – surpassing the ability of novice clinical decision-makers. Instead, the solutions rely on clinical context, and for that reason, the items overlap the practicum experiences of entry-level nurses. A panel of nursing SMEs validated all patient scenarios and items as entry-level appropriate (to be discussed more fully in the Item Review section below). They also verified that clinical judgments required taking into account multiple aspects of a client presentation.

Scenario	Clinical Judgment Element (Layer 3)	Expected Behavior(s) Adjust scenario/ item stems to provide information for each behavior.	Conditioning Factor(s) (Layer 4) Adjust scenario/ item stems to provide information for each factor.	Items Refer to the expected behaviors to help generate items. Mark the keys with an asterisk (*). Provide a rationale. Use resources as needed.	
	Recognize Cues	Recognize abnormal vs normal	Environment Cues:		
		Recognize signs and symptoms	Patient Observation Cues:		
		Identify history of	Medical Record Cues:		
			Time Pressure Cues:	Rationale	
				Internal Use Only Meets CJ step: Yes or No	Insert notes about accuracy, currency, fidelity, or entry-level here.
	Analyze Cues	Connection between pathophysiology and client presentation:	Requires knowledge of:		
		Use findings/ observations to determine client needs:		Rationale	
				Internal Use Only Meets CJ step: Yes or No	Insert notes about accuracy, currency, fidelity, or entry-level here.
	Prioritize Hypothesis	Prioritize (likelihood, risk, etc)	Requires knowledge of:		
			Indicate Resources:	Rationale	
				Internal Use Only Meets CJ step: Yes or No	Insert notes about accuracy, currency, fidelity, or entry-level here.
	Generate Solutions	Things to address:	Requires knowledge of:		
		Things to avoid:			
				Rationale	
				Internal Use Only Meets CJ step: Yes or No	Insert notes about accuracy, currency, fidelity, or entry-level here.
	Take Action	Request:	Requires knowledge of and experience with:		
		Administer:			
		Perform (Skill):			
		Document:			
		Communicate			
				Rationale	
				Internal Use Only Meets CJ step: Yes or No	Insert notes about accuracy, currency, fidelity, or entry-level here.
	Evaluate Outcomes	Reassess:	Requires knowledge of and experience with:		
				Rationale	
				Internal Use Only Meets CJ step: Yes or No	Insert notes about accuracy, currency, fidelity, or entry-level here.

Figure 3. Item writing task model template

2.3 Use of the Cognitive Continuum Theory in Task Model Development

The cognitive continuum theory postulates that decision-makers alternate between strategies as a function of the structure of the decision problem (Hammond, 1981; Harbison, 2001). Well-structured problems have a greater number of cues present that increase the propensity to rely on analytical decision-making options. By contrast, decision-makers rely on intuitive methods when the problem situation has fewer cues and greater levels of ambiguity.

In light of this theory, one focus of item development was to strike a balance between the reliance of these polar decision-making strategies. By providing enough structure for an entry-level nurse and not too much ambiguity, optimal decisions are reached through systematic and analytical methods. To ensure this, item writers added or removed structure to items by using exogenous and endogenous variables that are known to affect nursing clinical judgment (Muntean, 2012) layer 4 in Figure 2 and column 4 of Figure 3. The shaded elements in Figure 2 represent variables internal to the person (endogenous), and white background elements represent those external to the person (exogenous).

The task model template (Figure 3) was used for item development, where the shaded areas are the variables of interest for each clinical judgment element and the white boxes were for SMEs to fill in the appropriate information. Along with each clinical judgment element, there are two associated general factors of interest: expected behaviors and conditioning factors. The conditioning factors help item writers provide an elucidation of the main aspects of the clinical judgment element. For example, to recognize cues there are a number of opportunities to have cues revealed to the examinee such as observations of the patient or information found in a medical record, but also cues implicit to the severity of the presenting problem with respect to time pressure cues (i.e., what is the immediacy of the presenting problem). This structure helped to keep the process from becoming either overly structured or overly ambiguous.

3. Item Development Process

The item development process involved panels of representatively sampled nurse SMEs. A research team consisting of two master's level nurses with extensive clinical and educational experiences and two PhD psychometricians with a background in cognitive/clinical psychology facilitated the item development. The importance of having cross-functional facilitators stood out while teaching the construct to panelists and during panel discussions. Integrating expertise helped merge the cognitive nature of the construct into the context of the nursing domain.

The overall item development process had three major stages: item writing panels, item review panels, and item finalization. The item writing panels were responsible for creating entry-level scenarios, along with appropriate questions assessing the clinical judgment model. The item review panels were responsible for reviewing, providing feedback, and validating the scenarios and questions with respect to fidelity and accuracy. The item finalization stage incorporated the feedback and converted the base questions into the most appropriate response format (e.g., drop-down, hot-spot, drag-and-drop, etc.).

The item development process was a product of a number of iterations with different approaches. During each writing and review panel, facilitators took notes of successful and unsuccessful techniques. In addition, each panel concluded with a debriefing discussion soliciting ways to improve the process and identify things that worked well. The goal was to capture lessons learned and visions of how to implement productive changes for the next iteration of panels.

In many ways, the iterative approach that was taken with item development resembles the decision-making construct described in this paper. The panels were constructed with a goal in mind; however, information was gathered during the course of the item writing panel related to aspects of the process that went well and those that needed to be improved. The information was processed and analyzed in order to develop theories of explaining the observations. Solutions were then generated, prioritized, and implemented in an attempt to overcome issues and avoid impeding item writing in the next panel. Subsequently, the interventions were implemented in the following panel and the outcomes observed as to whether or not those solutions met the expectations. When they did not, the entire process was iterated again and new solutions identified. In the end, the following methodology yielded the most successful and fruitful item writing panel.

3.1 Item Writing Process

The item writing panels comprised between 6-8 SMEs. They began with an orientation of the clinical judgment construct. The facilitators went through definitions, practical examples, and previously developed items assessing the construct. In addition to the training, item writers had resource binders with the presented information and examples. This included definitions for each clinical judgment element, task statements and related content from the job analysis, and various writing guidelines and tips. This information was also presented on posters mounted across the room. Panelists understood that they could refer back to these resources during the item writing process.

After gaining an understanding of the construct, panelists were introduced to the task model template. This template was the basis for item production, and item writers filled in the appropriate areas of the template as they worked through scenario/item development. To orient the panelists to the task model and item development, the group of panelists led by the facilitators worked through an initial item as a group. This allowed for an opportunity for the panel to think through all aspects of the task model and get comfortable using the resources provided before working on their own. Additionally, throughout the panels, SMEs were encouraged to work together and discuss their scenarios and items with their colleagues.

The SMEs began filling out the task model by developing a clinical scenario appropriate for an entry-level nurse. Using that scenario, they abstracted all of the relevant information by filling in the variables in the Expected Behaviors and Conditioning Factors columns of the task model (column 3 & 4 of Figure 3). If information was missing from the narrative scenario, the SMEs edited it to ensure all required components were extant. The completed scenario served as the base information for items assessing each of the Clinical Judgment (CJ) elements.

Because the task model standardizes scenarios by abstracting informational units, the panelists were able to apply the template to produce a diverse set of clinical situations that were authentic, distinct, and varied. From these scenarios, panelists began writing items assessing each CJ element. Thus, all scenarios had six items written in an item set[1]. Items had several constraints: 1. each item must have between two and five correct answers; 2. distractors should resemble mistakes that SMEs had seen entry-level nurses do in the field; 3. items must have less than 10 options (keys plus distractors); and 4. Each item required a comprehensive and justifiable rationale. These questions and responses served as the raw material for building the final item sets.

SMEs were asked to write scenarios in terms of naturally occurring situations that entry-level nurses commonly encounter in their working environments. Because developing items related to specific clinical judgment elements was unusual to SMEs, they were asked to think through task model completion in a step-by-step process:

1. Develop a common clinical scenario that an entry-level nurse could be expected to encounter and would be appropriate for the entry-level
2. Describe or list the facts/observations and note the context within which the problem presents itself
3. Review the current information (shift handover reports, patient history reports, laboratory results, etc.) and gather new information where needed
4. Interpret the cues to understand the nature of the situation and the presenting signs or symptoms

    a. Distinguish between normal/abnormal
    b. Distinguish between relevant/irrelevant information
    c. Evaluate clusters of symptoms

5. Synthesize facts and make inferences to generate a hypothesis or a set of possible hypotheses of the patient problem and prioritize the set, e.g., more plausible to least, most important to address now, etc.
6. Formulate a course of action

    a. Identify a set of possible interventions to alleviate the presenting issue
    b. Select best alternative(s) from the set of possible options

7. Evaluate the effectiveness of the intervention, e.g. what would a positive/negative outcome look like, how would you know the intervention was effective, etc.
8. When writing distractors, think about commonly seen errors made by entry-level nurses relative to each of the CJ elements

The SMEs were also asked to think through the overall item development for a presenting scenario across all the CJ elements using the basic nursing process (Potter & Perry, 2009) as scaffolding:

1. Assessment: collect data about the patient and presenting problem
2. Nursing Diagnosis: organize the information from the assessment to make an informed judgment about the nursing diagnosis from the cluster of information
3. Outcomes/planning: based on the nursing diagnosis, delineate a set of measurable goals/outcomes and a course of treatment that moves from the present state to the expected future state
4. Implementation: carry out the planned care regime and monitor progress towards the expected outcome state
5. Evaluation: determine the effectiveness of the care plan

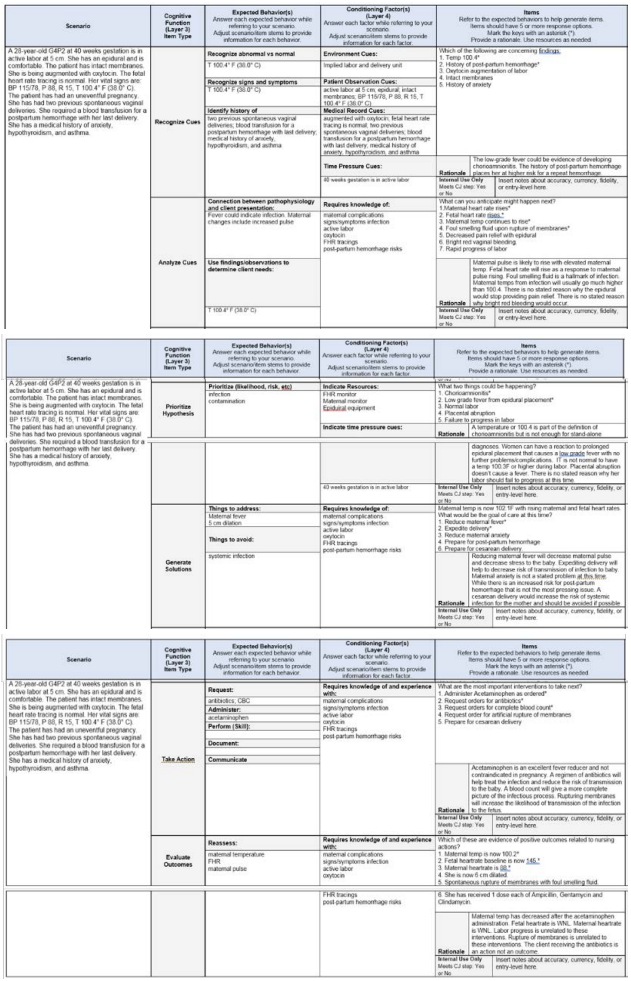

Figure 4. Example of a completed clinical judgment task model for item development.

The nursing process helped conceptualize the underlying cognitive elements defining clinical judgment with a well-known mental process in the nursing field. Figure 4 provides an example of a finished scenario and item set written by an SME during the item writing panel.

It is important to note that it took time for the SMEs to get into the swing of developing items in this manner. The main issues were that panelists were recruited with backgrounds that supported their expertise in their respective fields. Given what is known about experts (Benner, 1982) and was discussed above, experts tend to make decisions based on holistic thinking and have enough experience to organize and analyze situations intuitively. Therefore, it took time to help these experts take their scenarios and parse out the relevant clinical judgment elements from their scenarios. Having them reflect on the audience of the intended items, entry-level nurses, they were able to make the shift from expert perception to more specifically abstracting the cognitive components underlying the clinical judgment model in terms of the problem-solving aspects of the model. Also, providing the task model and some step-by-step procedures helped to provide structure to capture the specifics needed to measure CJ at the narrower level needed for evaluating entry-level.

For one example, the experts initially had a difficult time separating the recognition of cues from the analysis of those cues. As experts, they grasped the relevant and irrelevant cues intuitively because of their extensive experiences and analyzed them simultaneously. We asked them to think about the entry-level nurse and describe the plethora of cues that the individual would be seeing upon the initial presentation. Additionally, they were asked to place no value on the cues, but represent the actual cues that could be seen in that presenting situation and to think about cues that an entry-level nurse might think are relevant but are not. Then they were asked to develop an analysis item that showed the examinee could integrate the constellation of relevant cues to show that they were able to analyze them correctly. Additionally, we asked them to think about items that organized irrelevant cues or that an entry-level nurse might focus on and then ask what the most plausible analyses of those incorrect cues might be to help guide distractor writing. Thus, we attempted to take the holistic understanding of the expert SMEs and help them parse this into two distinct components using some basic heuristics.

In general, one SME could write at least four complete scenarios with the embedded six items over a two-day period. With greater experience, we expect SMEs to increase their writing capacity and generate a greater number of scenarios and clinical judgment questions.

During the debriefing sessions at the end of the writing panels, panelists would invariably say that they were just starting to get into the swing of this process at the end of the two-day period and felt as if the more they did this the easier it would become (E. Petersen & K. McMunn, personal communication, December, 12, 2018). This suggested that more consecutive writing days might have been useful and will be incorporated into future panels. Over time, it will be important to monitor the production of the panels to evaluate whether or not they are able to increase production with greater exposure and also to evaluate whether there are any changes in the underlying statistical functioning of items.

3.2 Item Review Process

When items are produced from the usual item writing panels that focus only on assessing the knowledge component of the profession, it is necessary to review the items for accuracy, currency, and reference the correct key with professional literature. Moreover, for the current project, reviews were needed to verify that individual items written to each CJ element were truly measuring that element. This was vitally important as the validity of the score based on those items is paramount, and there is no easy reference to consult to ensure an item is measuring a specific CJ element. However, references can be found to validate the individual correct answer keys for each item written within each CJ element. Therefore, the item review panels followed a slightly different path in order to handle this new aspect of item development. The process for reviewing first focused on reviewing the CJ scenario's narrative presentation. Then a review of the items was undertaken as a second phase. As with the writing panels, a number of iterations were undertaken that resulted in the process reported here.

When reviewing the CJ narrative scenarios, the SMEs on the review panel needed to answer a number of questions. First, is the scenario something that an entry-level nurse would see in actual clinical practice? Is the presentation of the scenario authentic? Does all the information provided seem realistic? It was important at this point to ensure that there was fidelity to the scenarios and also assure that it was appropriate for an entry-level nurse in the late beginner stage of the skills acquisition levels. The generic overview was: is this a realistic scenario that an entry-level nurse would experience and would be expected to be able to handle?

Next review would turn to each of the items written for each of the CJ elements. Reviewers were asked to review the items as they normally would for accuracy, currency, and correctness of answer keys. However, they were also asked to validate both the rationale given by the item writer and the information provided in the middle columns of the template to ensure accuracy. The reviewers would need to answer a number of questions, such as: Is the item written coherently? Are the answers accurate? Is the practice current? Is the question appropriate for the entry-level nurse? Does the rationale provide validation for the correct options and rule out distractors? Are the contextual factors and expected behaviors documented correctly? Does the item measure the intended CJ element? Is the item authentic and appropriate for the context of the scenario?

In addition, another set of review questions was used when reviewing item sets comprising a scenario. As the item sets are related to a common patient and issue, it was important to evaluate the consistency of information across sequential items. To this end, reviewers were asked to consider each question and determine if the information presented in the scenario was needed to answer the question. If not, then the item was deemed a knowledge item and not expected to measure the intended CJ element.

Next, reviewers were asked to evaluate the consistency of option content across items. This was deemed necessary to ensure that cueing across items would not happen as a result of different sets of distractor options. For instance, if a subsequent item did not have distractor options that were reasonable options based on the distractor options from the prior items, it was felt that this could potentially cue examinees to the correct answers. Moreover, having the consistency of distractor options was deemed important in order to evaluate whether or not individuals would be able to recognize with new information that they had initially gone down an incorrect path but could self-correct, which is an important aspect of competent decision-making.

This process was meant to provide a comprehensive method for providing evidence for the legitimacy of each item and the item set as a whole. Additionally, it was felt that this would provide defensibility in lieu of a simple, single reference in a textbook for the entire scenario. It should be noted that documented references were used to justify information within the rationale and for each

individual component of the expected behaviors and conditioning factors.

In general, a panel of six SMEs was able to review about 180 items along with their related scenarios over 2.5 days. Also, roughly 75% of the items were deemed to be useful and need only minimal revisions or editorial work to be deemed ready to test operationally, another 20% were deemed in need of significant revision, and about 5% were deemed to be significantly poor that they were removed from further consideration (K. McMunn & E. Petersen, personal communication, December, 12, 2018).

4. Final Item Development

The resulting scenarios and item sets developed from the writing panels and validated by the review panels were next developed into final items using a number of different item/response types. The results of the writing and review panels were taken as the raw materials from which final items would be designed using a number of validated[2] item/response designs that are specific instantiations of commonly used technology-enhanced item/response types (Parshall, Davey, & Pashley, 2000; Parshall & Harmes, 2007; Scalise & Gifford, 2006; Sireci & Zenisky, 2006). The work of the test development SMEs was to find the appropriate design for rendering the final items using the following item design and response types.

4.1 Item/Response Type Models

Multiple Response (MR). This response type provides the examinee with a series of options associated with check boxes which allow for multiple correct keys to be evaluated in a single item. It also allows for the potential for over- or under-responding by the candidate as the items are written to 'select ALL' that apply. No direction is given concerning the correct number of keys for an item (example seen in Figure 7).

Hot-spot/Highlighting (HL). These types of items allow for sections of text or graphics to be tokenized such that when the examinee the cursor or tabs across elements of the response field to highlight that section. The candidate only needs to click on the highlighted area or press the enter button to select that option/token as a response. As with the MR items, these items can be presented with a 'select ALL' stem allowing for over- or under-responding.

Two approaches were taken to the design of this type of response method. The first approach was to provide textual information which is a common method for conveying information in the nursing field. Then phrases were tokenized, and candidates were able to click on as many of the phrases as they deemed appropriate. An example of this is Figure 5 where the text in the nurses' notes has phrases tokenized for which the candidate can select.

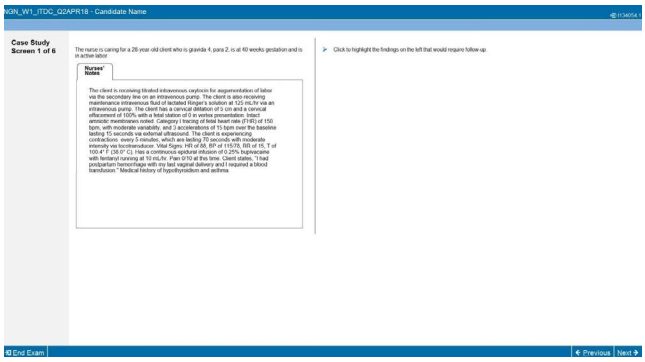

Figure 5. Final recognize cues item from Figure 4.

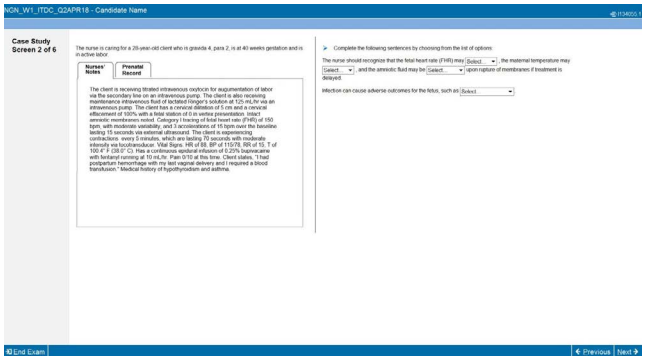

Figure 6. Final analyze cues item from Figure 4.

The second approach to using this response structure was when information was presented in a tabled format. An example of this could be a laboratory report or a schedule of medications, or a table tracking important information about a patient over time. In this case, the candidates were expected to highlight the correct row(s). The table type items were not restricted to only a single column and could span numerous columns if needed. This response structure was similar to the MR items above as a 'select all' direction was given.

Drop-down (DD). This type of item used the standard DD type menu. There were two variations on the use of this response structure. The first was when the DD element was embedded within a table structure. In this format, cells of information within the table rows/columns were represented as items where pieces of important information needed to be selected to complete the clinical picture

of the presenting situation. The other approach was to embed the DD structure within the context of a continuous text, e.g., similar to a Cloze procedure in reading comprehension studies. An example of this type of use of the DD can be seen in Figure 6 where items are embedded within a text such that the candidate must select the correct responses to complete the analysis of cues.

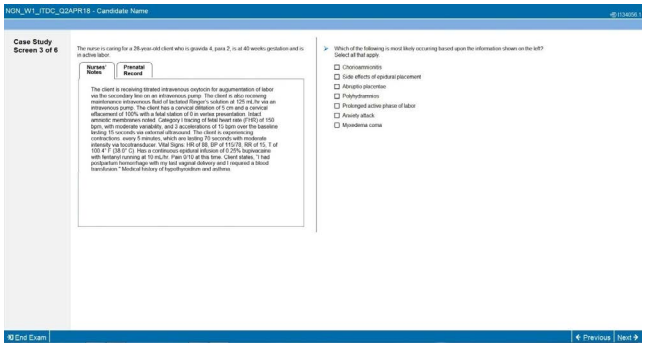

Figure 7. Final prioritize hypothesis item from Figure 4.

Drag-and-Drop (DND). This type of item provides for a set of token elements to be enabled to be dragged to specific target locations to answer the question. The draggable elements are called tokens, and the response area is called the target. Examples of these types of items can be found in Figures 8 and 9. Figure 9 provides the usual layout of the item where there are a number of tokens that need to be moved to the appropriate targets. For these items there were two variations on the presentation which corresponded to directions that stated either 'select all' or 'select N' that apply. Figure 9 is an example of a 'select all' which allows for candidates to over- or under-respond. In a situation where a specific number of responses were required (e.g., 'select 3'), then only the appropriate number of targets were presented (e.g., 3 targets), respectively.

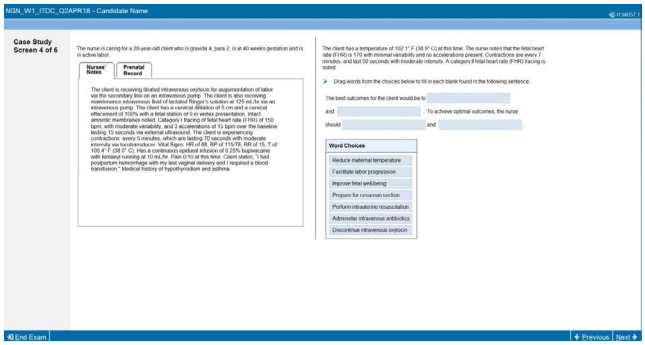

Figure 8. Final generate solutions item from Figure 4.

This type of response structure also can be used when ranking or priority type items are needed. In this case, a column is added to the target area that provides the labels for each of the rows. For instance, a target can be related to the highest risk symptom or most immediate intervention to take. Additionally, DND can be used in a matching type response structure that allows for evaluating clinical judgment for tightly coupled aspects of the decision-making process. For example, in some situations you want to take a specific action to respond to a specific symptom, using a matching approach the item can be designed such that tokens A and B might be a tightly coupled pair that should be placed next to each other in the same row under two separate column headings.

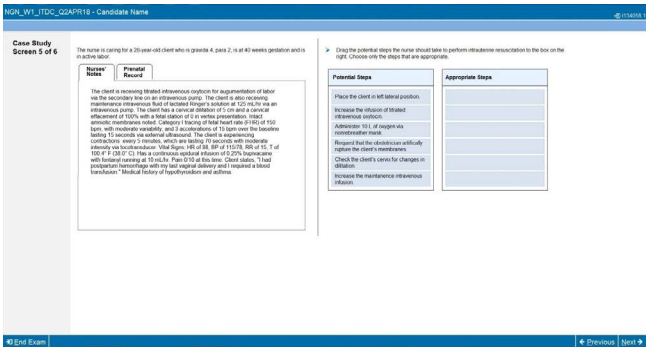

Figure 9. Final take action item from Figure 4.

Another type of DND is shown in Figure 8. This is a variant of the DD Cloze type item where the DD elements are replaced with targets. The candidate selects the correct token from the choice options and drags them to the appropriate target. This item allows for the tokens to be dropped on the target areas that are within the sentences. This item type was developed based on feedback from the usability and cognitive labs. The main issue with the DD approach to this type of item was that all of the information was occluded when wrapped in a DD item response type. Therefore, the participants felt it was more complicated and time-consuming to open all the DD items and review the options; likewise, when the focus was on one DD item, the view of the elements in the other DD responses would close. This item type was an attempt to enable the participants to view all answer choices in a single viewing and potentially reduce the cognitive and emotional issues raised with the DD presentation.

Matrix/Grid. This response type can be seen in Figure 10. This type of response comes in two types, multiple choices and multiple responses. Figure 10 shows the multiple choices variant. For this variant, each row is

conditionally independent of the others, and there is one correct option per row. This presentation is useful when presenting a number of assessment findings or reactions to interventions and then asking whether each of those indicators is in one of several independent sets. For instance, Figure 10 has three independent sets for which a number of assessment findings are indicative of three outcomes: an effective intervention, an ineffective intervention, or is unrelated to the intervention.

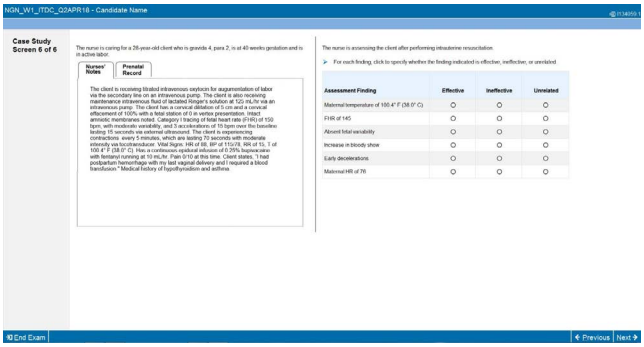

Figure 10. Final evaluate outcomes item from Figure 4.

When the multiple response variant is used, then each column becomes the independent item. This type of presentation is useful when there are a number of symptoms defining the rows and the columns represent possible disease processes. This allows for presentation of numerous symptoms from a case study and asking candidates to identify which sets of symptoms are indicative of which disease process when the symptoms are not necessarily independent across disease processes.

4.2 Development of Final Item: Combining Raw Material from Panels and Item/Response Types

The item/response types described above were used to develop individual items related to the outcome of the item writing panel. Figure 4 showed the raw material that was generated during a writing panel. SMEs and content developers then took this information and used the above item/response types to build a set of final, ready for presentation item sets. Note that the development of the items from the task model entails writing one item for each of the six clinical judgment elements, and therefore all of the items are related to a single patient scenario that measures each of the elements of the clinical judgment model.

Figures 5-10 provide the finished product based on the raw item data displayed in Figure 4. This set of six items is called the scenario and defined as the collection of items related to the common presenting situation that steps through all of the clinical judgment model elements. Each item measures one element of the CJ model. The design of the items uses a left side, shaded border with an indicator of the current screen number in the set of screens for the scenario. The nomenclature of "Case Study" is used for the scenario because this is a common way of presenting client situations in the nursing field. Each of the items was designed to faithfully represent the content of the item writer's intentions and validated by the item review panel but incorporate the appropriate response type for gathering data. It was common that a number of different response types were used across the set of items pertaining to a scenario.

This specific scenario progresses from a HL type item that asks the examinee to recognize the important cues in the client presentation and then uses a Cloze type DD item to evaluate how successfully the examinee can analyze the clues presented. Also notice that new information has been provided when moving from the 1st to the 2nd item in this scenario in "Prenatal Record" tab. Then a MR response type is used to gather information about a number of possible hypotheses that would result from the analysis. The next item then provides updated information in the upper, right side of the item and uses the DND Cloze type item to now identify whether or not the examinee can generate appropriate solutions given the updated situation. The next item then allows the examinee to select from a number of actions to take in order to address the solutions. Finally, a matrix/grid item is used to evaluate whether or not the examinee is able to evaluate the effectiveness of intervention given a set of plausible outcomes. Scoring will be discussed below.

The final items present a significant amount of on-screen information compared with the usually, relatively bare bones multiple choice type item. However, all information from previous items will carry over to subsequent items. So while the items are text-heavy, one does not need to completely read each item anew, but only read the new additional information when provided or review previously read text.

5. Discussion

Establishing an accurate measure of higher-order cognitive constructs like clinical judgment is a challenging task.

But the process is made easier when broken down into smaller, manageable objectives. This study establishes a methodology for one component – item development – that fits into the larger scope of measurement and can be generalized to other fields. One of the core components was the iterative approach where future endeavors were based on continually critiquing what worked and what did not work in previous research endeavors. Additionally, the task model provides a systematic approach to generating scenarios and questions that bridge the gap between theory and practice. From there, rendered items best represent their intended construct.

Using a representative sample of SMEs allowed us to capture the diverse nature of clinical judgment that accurately resembles the state of the nursing field. However, the more diverse the practice, the more difficult it is to capture the unique idiosyncrasies. The task model solves this challenge by standardizing item writing so that SMEs from all backgrounds can develop unified scenarios and questions. However, the process of implementing the task model could be difficult at first as almost all of the SMEs used for the writing were themselves experts and had a tough time abstracting their scenarios into the more specific pieces that make up the CJ model. After continued discussion and facilitation, SMEs were able to more clearly separate their ideas into the requisite CJ model components.

Through instruction, SMEs successfully wrote to the appropriate responsibility level of entry-level nurses, as confirmed by the review panel. Hosting an independent review panel generated valuable feedback on both the written items and on the item development process. Through an iterative process, items were rendered into their final form by reworking the raw content from the task model template into response formats that lend themselves to the objective of the question. In the end, the item development process is central to accurate measurement, and our methodology sets a strong foundation for measuring higher-order cognitive constructs.

The evaluation of clinical judgment or context dependent decision-making is vital to many professions outside of nursing. Evaluating examinees level of decision-making can help define competent practice across a host of professions that goes beyond simple knowledge of facts. The process presented here could provide a launching point for other test development programs to begin the investigation of evaluating examinees competency in skilled decision-making. Overall, the approach presented here can be generalized to other programs that would like to measure the complex cognitive components of their professional domain.

6. Conflict of Interest

There are no conflicts of interest related to publication of this article. All research documented within this manuscript were part of the normal R&D process that guides test development. The research was sponsored by the National Council of State Boards of Nursing (NCSBN). All authors work to support the on-going operational and R&D work for the NCLEX® examinations owned by NCSBN and delivered by Pearson VUE.

7. Acknowledgment

The authors would like to thank Emily Petersen, MS, MJ, RN, APRN, CPNP and Kelly McMunn, MSN, RN, CMSRN for their assistance in this research. Their experience and subject matter expertise was instrumental in facilitating item writing and review panels, and was an important key for success.

8. References

Benner, P. (1982). From novice to expert. *The American Journal of Nursing, 82(3)*, 402–07. https://doi.org/10.1097/00000446-198282030-00004. https://doi.org/10.2307/3462928. PMid: 6917683.

Bereiter, S. R. and Miller, S. M. (1989). A field-based study of troubleshooting in computer-controlled manufacturing systems. *IEEE transactions on Systems, Man, and Cybernetics, 19(2)*, 205–19. https://doi.org/10.1109/21.31027.

Berkow, S., Virkstis, K., Stewart, J. and Conway, L. (2008). Assessing new graduate nurse performance. *Journal of Nursing Administration, 38(11),* 468–74. https://doi.org/10.1097/01.NNA.0000339477.50219.06. PMid: 18997551.

Berman, A., Snyder, S.J. and Frandsen, G. (2016). *Kozier and Erb's fundamental of nursing: concepts, process and practice* (10th ed.). Upper Saddle River, NJ: Pearson Education, Inc. https://www.pearson.com/us/higher-education/product/Berman-Kozier-Erb-s-Fundamentals-of-Nursing-10th-Edition/9780133974362.html.

Berman, A. and Snyder, S. (2012). *Kozier and Erb's Fundamentals of Nursing: Concepts, process, and practice* (9th ed.). Pearson. https://www.pearson.com/us/higher-education/product/Berman-Kozier-Erb-s-Fundamentals-of-Nursing-9th-Edition/9780138024611.html.

Burton, M.A. and Nay Ludwig, L.J. (2015). *Fundamentals of nursing care: Concepts, connections and skills*, (2nd ed.). Davis. ISBN-13: 978-0-8036-3974-4. https://www.fadavis.com/product/nursing-fundamentals-med-surg-nursing-care-concepts-connections-skills-2-burton-ludwig.

Casey, K., Fink, R. R., Krugman, A. M. and Propst, F. J. (2004). The graduate nurse experience. *Journal of Nursing Administration, 34(6)*, 303–11. https://doi.org/10.1097/00005110-200406000-00010. PMid: 15190226.

Craven, R.F., Hirnle, C.J. and Henshaw, C.M. (2017). *Fundamentals of nursing: Human health and function* (8th ed.). Philadelphia: Lippincott, Williams & Wilkins. ISBN-13: 978-1469898605.

Dickison, P., Haerling, K., and Lasater, K. (2019). Integrating the National Council of State Boards of Nursing clinical judgement model into nursing. *Journal of Nursing Education,* 58(2), 72-78. https://doi.org/10.3928/01484834-20190122-03

Dickison, P., Luo, X., Kim, D., Woo, A., Muntean, W. and Bergstrom, B. (2016). Assessing higher-order cognitive constructs by using an information-processing framework. *Journal of Applied Testing Technology, 17(1)*, 1–19.

Ebright, P. R., Urden, L., Patterson, E. and Chalko, B. (2004). Themes surrounding novice nurse near-miss and adverse-event situations. *Journal of Nursing Administration, 34(11)*, 531–38. https://doi.org/10.1097/00005110-200411000-00010. PMid:15586075.

Elstein, A. S., Shulman, L. S. and Sprafka, S. A. (1978). *Medical problem solving: An analysis of clinical reasoning.* Cambridge, MA: Harvard University Press. https://doi.org/10.4159/harvard.9780674189089.

Embretson, S. (1983). Construct validity: Construct representation versus nomothetic span. *Psychological Bulletin, 93,* 179–97. https://doi.org/10.1037/0033-2909.93.1.179.

Hammond, K. (1981). *Principles of organization in intuitive and analytical cognition* (Report 231). Center for Research on Judgment and Policy, University of Colorado, Boulder, CO. p. 80. https://apps.dtic.mil/dtic/tr/fulltext/u2/a096570.pdf.

Harbison, J. (2001). Clinical decision making in nursing: theoretical perspectives and their relevance to practice. *Journal of Advanced Nursing, 35(1),* 126–33. https://doi.org/10.1046/j.1365-2648.2001.01816.x. PMid: 11442690.

Hickey, M. T. (2009). Preceptor perceptions of new graduate nurse readiness for practice. *Journal for Nurses in Professional Development, 25(1)*, 35–41. https://doi.org/10.1097/NND.0b013e318194b5bb. PMid:19182556.

Ismail, Z. and Trotman, K. T. (1995). The impact of the review process in hypothesis generation tasks. *Accounting, Organizations and Society, 20(5)*, 345–57. https://doi.org/10.1016/0361-3682(95)00002-Q.

Kavanaugh, J. M. and Szweda, C. (2017). A crisis in competency: The strategic and ethical imperative to assessing new graduate nurses' clinical reasoning. *Nursing Education Perspectives, 38(2),* 57–62. https://doi.org/10.1097/01.NEP.0000000000000112. PMid: 29194297.

Kuiper, R., O'Donnell, S. M., Pesut, D. J. and Turrise, S. L. (2017). *The Essentials of Clinical Reasoning for Nurses: Using the Outcome-Present State Test Model for Reflective Practice.* Sigma Theta Tau. ISBN-13: 978-1945157097.

Lasater, K. (2007). Clinical judgment development: Using simulation to create an assessment rubric. *Journal of nursing education, 46(11),* 496–503. PMid: 18019107.

Lasater, K., Nielsen, A. E., Stock, M. and Ostrogorsky, T. L. (2015). Evaluating the clinical judgment of newly hired staff nurses. *The Journal of Continuing Education in Nursing, 46(12),* 563–71. https://doi.org/10.3928/00220124-20151112-09. PMid: 26641154.

Libby, R. (2017). *Accounting and human information processing. In: The Routledge Companion to Behavioural Accounting Research,* Routledge. p. 42–54.

Masters, G. N. (2016). Partial credit model. In: *Handbook of Item Response Theory, Volume One.* Chapman and Hall/CRC. p. 137–54.

Muntean, W. (2012). *Nursing clinical decision-making: A literature review.* Retrieved at https://www.ncsbn.org/Nursing_Clinical_Decision_Making_A_Literature_Review.htm.

National Council of State Boards of Nursing (2018). *Strategic Practice Analysis.* NCSBN Research Brief, 71. Chicago, IL: NCSBN. https://www.ncsbn.org/11995.htm.

Nibbelink, C. W. and Brewer, B. B. (2018). Decision-making in nursing practice: An integrative literature review. *Journal of Clinical Nursing, 27(5-6),* 917–28. https://doi.org/10.1111/jocn.14151. PMid: 29098746, PMCid: PMC5867219.

Nichols, P. D., Kobrin, J. L., Lai, E. and Koepfler, J. (2016). The role of theories of learning and cognition in assessment design and development. *The handbook of cognition and assessment: Frameworks, methodologies, and applications,* 15–40. https://doi.org/10.1002/9781118956588.ch2.

Oppenheimer, D. M. and Kelso, E. (2015). Information processing as a paradigm for decision making. *Annual Review of Psychology, 66*, 277–94. https://doi.org/10.1146/annurev-psych-010814-015148. PMid: 25559114.

Parshall, C. G., Davey, T. and Pashley, P. (2000). Innovative item types for computerized testing. In: W.J. van der Linden and C.A.W. Glas (eds.), *Computerized Adaptive Testing: Theory and Practice.* Netherlands: Kluwer Academic Publishers. https://doi.org/10.1007/0-306-47531-6_7. PMid: 10947477.

Parshall. C. G. and Harmes, J. C. (2007). Designing templates based on a taxonomy of innovative items. In: D. J. Weiss (Ed.). *Proceedings of the 2007 GMAC Conference on Computerized Adaptive Testing.* Retrieved from www.psych.umn.edu/psylabs/CATCentral/.

Potter, P. and Perry, A. (2009). *Fundamentals of Nursing* (7th ed). Mosby Elsevier.

Potter, P.A., Perry, A.G., Stockert, P.A. and Hall, A.M. (2017). *Fundamentals of nursing* (9th ed.). St. Louis: Elsevier.

Raymond, M. (2016). Job analysis, practice analysis and the content of credentialing examinations. In S. Lane, M. Raymond, and T. Haladyna (Eds), *Handbook of test development* (2nd ed.), Routledge. p. 144–64.

Rubenfeld, G. and Scheffer, B. (2010). *Critical thinking tactics for nurses: Achieving the IOM competencies.* Jones & Bartlett Learning. https://doi.org/10.1016/S2155-8256(15)30371-9.

Saintsing, D., Gibson, L. M. and Pennington, A. W. (2011). The novice nurse and clinical decision-making: How to avoid errors. *Journal of Nursing Management, 19(3),* 354–59. https://doi.org/10.1111/j.1365-2834.2011.01248.x. PMid: 21507106.

Scalise, K. and Gifford, B. (2006). Computer-Based Assessment in E-Learning: A Framework for Constructing «Intermediate Constraint» Questions and Tasks for Technology Platforms. *Journal of Technology, Learning, and Assessment, 4(6)*. Retrieved from http://www.jtla.org.

Schaper, N. (1998). Analysis and training of diagnostic expertise in complex technical domains. *European Journal of Work and Organizational Psychology, 7(4),* 479–98. https://doi.org/10.1080/135943298398529.

Simmons, B., Lanuza, D., Fonteyn, M., Hicks, F. and Holm, K. (2003). Clinical reasoning in experienced nurses. *Western Journal of Nursing Research, 25(6),* 701–19. https://doi.org/10.1177/0193945903253092. PMid: 14528618.

Sireci, S. and Zenisky, A. (2006). Innovative item formats in computer-based testing: In pursuit of improved construct representation. In: S. M. Downing & T. M. Haladyna (Eds.), *Handbook of test development,* Mahwah, NJ: Lawrence Erlbaum Associates. p. 329–347.

Standing, M. (2014). *Clinical judgment and decision making for nursing students* (2nd ed). Sage.

Su, Y. L. and Govindaraj, T. (1986). Fault diagnosis in a large dynamic system: Experiments on a training simulator. *IEEE Transactions on Systems, Man, and Cybernetics, 16(1),* 129–41. https://doi.org/10.1109/TSMC.1986.289289.

Tanner, C. A. (2006). Thinking like a nurse: A research-based model of clinical Judgement in nursing. *Journal of nursing education, 45(6)*, 204–11. PMid: 16780008.

Taylor, C.R., Lillis, C., Lynn, P. LeMone, P. (2015). *Fundamentals of nursing: The art and science of nursing care* (8th ed.). Philadelphia: Lippincott, Williams & Wilkins.

Thompson, C. and Dowding, D. (2009). *Essential Decision Making and Clinical Judgement for Nurses E-Book*. Elsevier Health Sciences.

Wilkinson, J. (2012). *Nursing process and critical thinking* (5th ed). Pearson.

Wilkinson, J.M., Treas, L.S., Barnett, K.L. and Smith, M.H. (2016). *Fundamentals of nursing* (Vol. 1-2, 3rd ed.). Philadelphia: F.A. Davis Company.

Notes

[1] Although the panelists created item sets assessing each element, the task model template can be used to develop items measuring individual clinical judgment elements without being part of the scenario item set.

[2] Usability studies were previously employed to provide evidence in support of the final item/response designs.