

Situational judgment test validity: An exploratory model of the participant response process using cognitive and think-aloud interviews

Michael D. Wolcott (✉ wolcottm@email.unc.edu)

University of North Carolina at Chapel Hill <https://orcid.org/0000-0003-3254-3643>

Nikki G. Lobczowski

Carnegie Mellon University

Jacqueline M. Zeeman

University of North Carolina at Chapel Hill

Jacqueline E. McLaughlin

University of North Carolina at Chapel Hill

Research article

Keywords: cognitive interview, empathy, qualitative methodology, response process, situational judgment test, think-aloud protocol, validity

Posted Date: April 1st, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-19804/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Version of Record: A version of this preprint was published on December 14th, 2020. See the published version at <https://doi.org/10.1186/s12909-020-02410-z>.

Abstract

Background: Situational judgment tests (SJTs) are used in health sciences education to measure knowledge using case-based scenarios. Despite their popularity, there is a significant gap in the validity evidence and research on the response process that demonstrate how SJTs measure their intended constructs. Models of the SJT response process have been proposed in the literature; however, few studies explore and expand these models beyond surface-level attributes. The purpose of this study was to describe the factors and strategies involved in the cognitive process examinees use as they respond to SJT items.

Methods: Thirty participants—15 students and 15 experienced practitioners—completed a 12-item SJT designed to measure empathy. Each participant engaged in a think-aloud interview while completing the SJT followed by a cognitive interview probing their decision-making processes. Interviews were transcribed and independently coded by three researchers to identify salient themes and factors that contributed to the response process.

Results: Results suggested that the SJT response process included the complex integration of comprehension, retrieval, judgment, and response selections. Each of these response process stages were influenced by attributes such as perceived objective of the task, job-specific knowledge, assumptions about the scenario, and item setting.

Conclusions: This study provides an evaluation of the SJT response process and contributes exploratory information to the validity evidence of SJTs; these findings can inform the design, interpretation, and utility of SJTs.

Background

Situational judgment tests (SJT) have attracted substantial interest in health sciences education as an assessment methodology.^{1,2} SJTs are designed to evaluate how an examinee would respond to scenarios commonly encountered in practice.^{3,4} During an SJT, the examinee is presented with a hypothetical scenario and asked to evaluate the effectiveness of potential responses to that scenario. SJT items are designed to capture the examinee knowledge by selecting the most appropriate response consistent with job expectations, which are also tied to the constructs being measured.⁵ Participants are then assigned a score based on how well their selections align with a key, often established using subject matter experts.⁶ High scores often indicate an individual with high levels of the trait being evaluated (i.e., knowledge pertaining to the construct or constructs of interest).

SJTs have been integrated into admissions processes, capstones, and longitudinal assessments across various disciplines including medicine, pharmacy, and nursing.^{2,7-10} Despite increasing popularity, research on the SJT methodology is particularly arduous due to the rapidly changing foci about what matters in the field as well as the evolving theoretical and empirical understanding of SJTs.¹¹ Escalating

interest in SJTs initially eclipsed efforts to generate supporting evidence for its use as an assessment strategy in health sciences education. Specifically, there were inadequate attempts to establish validity evidence that distinguished what constructs were assessed and the elements involved in response processes.¹² It is imperative that assessments have sufficient validity evidence to support their interpretation and use.¹³ Of the five sources of validity evidence recommended by the Standards for Educational and Psychological Testing, research on the response process during SJTs is a neglected area of research despite clear deficits on this topic according to SJT researchers.^{12,14-20}

This limitation of research restricts our understanding of the cognitive processes involved when completing SJT items. Considering an SJT is intended to measure a decision-making or judgmental process, there are various response processes that are assumed to take place that have not been thoroughly described in the literature. During any assessment, an examinee activates a cognitive response process, which includes the moment-to-moment steps required to think and make decisions.²¹ Cognitive response processes, therefore, include how information is accessed, represented, revised, acquired, and stored to address a question. The decision-making processes include the manipulation of information in a series of steps, which can be informed by existing knowledge, experience with previous techniques, or the application of analogies; this process is also triggered by contextual cues. In general, cognitive response processes associated with schema that are domain-specific and, therefore, change depending on the setting.²¹

When it comes to assessing complex thought processes for assessments, evidence must demonstrate that test takers use cognitive processes in a coordinated fashion that is consistent with the theoretical and empirical expectations.²² Evaluating cognitive response processes is often elaborate and can vary based on the context or the tasks being assessed. Evidentiary sources investigating cognitive response processes often include think-aloud procedures and cognitive interviews, which are used as part of an overall cognitive task analysis to create verbal reports that can be annotated and analyzed.^{23,24}

A foundational perspective of assessing cognitive processes refers to research on cognitive aspects of survey methodology,²⁵ which is also applicable to assessing SJT response processes because they both involve situating oneself in the context and choosing responses that would be guided by schema relevant in those situations. This approach considers the task characteristics and respondent behaviors to describe the interplay between cognitive and communicative processes necessary for response to survey items. Tourangeau, Rips, and Rasinski proposed a four-step process that participants use when completing a survey: (1) comprehension, (2) retrieval, (3) judgment, and (4) response selection.²⁶ During comprehension, the examinee uses cognitive processes to read, interpret, and understand the purpose of the question. Next, the retrieval phase includes accessing long-term memories and knowledge relevant to the scenario and proposed problem. A judgment is formed by the examinee based on a complex integration of memories, knowledge, experiences, and other antecedents.²⁷ Finally, the examinee selects a response that is most consistent with their judgments.

Ployhart proposed an SJT response model that added contextual factors specific to an SJT using the four-stages proposed by Tourangeau and colleagues as the foundation of this process.^{18,26} In general, he noted that sources of construct-irrelevant variance (e.g., language barriers, interpretation issues, and impression management) can affect all stages of the response process, in addition to overall test-taking motivation.¹⁸ Since there is little known about the response processes governing how individuals interact and respond to SJT items,¹² the following study was conducted to evaluate whether these elements are salient in the cognitive processes engaged when completing an SJT.

The purpose of this study was to identify the salient elements of the SJT response process, thus addressing an important gap in the validity evidence literature. This study focused on generating evidence of response processes to an SJT measuring an important construct in health professions training: empathy. This research provides a prototype for exploring and describing response processes when using construct-focused SJTs with the intent of applying the methodology to SJTs measuring other constructs of interest in the future. The research question to address was: What factors and strategies are involved in the cognitive processes when examinees respond to SJT items? Due to the limited evidence regarding the response processes when completing an SJT, the research question was exploratory in nature. The goal was to probe examinees during an SJT to better understand what factors impacted the decision-making and response process using think-aloud and cognitive interviews.

Understanding of SJT response processes is critical because SJTs are assumed to engage cognitive processes related to decision-making abilities and prioritization of actions, which has not been demonstrated empirically. Knowledge of these processes can inform SJT design by identifying how design elements impact the response as well as awareness of individual attributes that may introduce construct-irrelevant variance into the score.

Methods

Participants

Due to the exploratory nature of the study and the necessity to conduct in-depth interviews with participants, the study included a convenience sample of fifteen student pharmacists enrolled in a Doctor of Pharmacy (PharmD) degree program and 15 pharmacy practitioners with at least 5 years of experience. The sample size was deemed sufficient based on prior SJT response process research.¹⁹ Participants were provided an alphanumeric identifier to designate if they were a student participant (indicated by the label “S” followed by a number from 1 to 15) or a pharmacist participant (indicated by the label “P” followed by a number from 1 to 15). The study was approved by the University of North Carolina Institutional Review Board.

SJT Specifications

An SJT was created to evaluate empathy (i.e., the construct of interest) given its multifaceted nature and relevance to healthcare.^{28,29} Empathy is considered a multidimensional construct that includes at least two factors: cognitive empathy and affective empathy.³⁰⁻³³ *Cognitive empathy* refers to an individual's ability to understand another person's perspective versus being self-oriented.³⁴ This cognitive perspective includes being able to imagine alternative realities, to judge the difficulty of scenarios, and to "step into another person's shoes and to step back as easily into one's own shoes again when needed".³¹ The other element, *affective empathy*, pertains to an individual's ability to understand and internalize the feelings experienced by others.³⁵ Also called emotional empathy, affective empathy relates to recognizing the emotional response that can be generated by individuals or through the interactions between people.³¹

The design of SJT items for this study was based on Lievens' construct-driven approach, which incorporates theoretical and empirical evidence to inform sound instrument design.³⁶ Each item was designed to target one of the two components of empathy (i.e., cognitive or affective empathy). The overall score on the SJT was, therefore, representative of the unidimensional construct of empathy. The intended inference was that high scores on the SJT (i.e., examinee answers are most consistent with the keyed answers) were indicative of exhibiting more empathy, whereas low scores suggested a lower degree of empathy. The SJT was designed to include items that varied based on empathy component (i.e., affective and cognitive empathy) and setting (i.e., healthcare and non-healthcare environments).

SJT items were structured using a knowledge-based format (i.e., *should do*), as this format has been shown to require more job-specific and general knowledge.^{37,38} All items used ranking-response formats, as this requires participants to analyze and discriminate among all options for each test item.^{39,40} To allow the participants ample time to answer each question, their response time was not restricted, however, it was estimated that participants would require at least two minutes per question.

SJT Item Development

All items were developed and reviewed by subject matter experts, which consisted of 11 faculty and pharmacy practitioners (e.g., experts) from multiple practice settings (e.g., research, ambulatory care, community, and hospital settings). The item developers were predominantly female (n=6, 54%) with an average of 13.5 years practice experience. Individuals were divided into two groups: seven functioned as item writers and four served as item reviewers—the two-step process allowed for extensive review of the items before their use in the study.

The assigned item writers created 24 items with instructions to create items that targeted affective and cognitive empathy in both healthcare and non-healthcare settings. Attention was paid to the content and response options for each item as this SJT was intended to reflect plausible scenarios. Item designers utilized information from published literature, practice analyses, personal experiences, sample SJT items from the health professions, and theoretical constructs to guide the development of each item. The item designers worked in small teams to craft the items and create a draft key for the response ranking.

The four item reviewers independently evaluated all 24 items on three criteria: (1) how well the item measured empathy, (2) the component of empathy assessed, and (3) the type of setting used in the item. These experts were also required to independently rank the response options to determine the final key for the item, which served as an additional evaluation criterion (e.g., the agreement in their answer keys). SJT items were included in the final test if there was a high level of agreement among subject matter experts on the ranking of response options (e.g., Kendall's coefficient of concordance above .6),^{41,42} if the item was perceived to be a good measure of empathy, and if there was majority agreement (e.g., at least 3 of the 4 reviewers) identified the specific component of empathy and the setting (i.e., cognitive or affective empathy in a healthcare or non-healthcare setting). A fifth reviewer was recruited to break ties where necessary.

The final SJT used in the study included 12 items with 6 items per empathy component. Three items of each component targeted job-specific knowledge (i.e., a healthcare setting) and three items targeted general domain knowledge (i.e., a non-healthcare setting). Figure 1 provides a visual representation of the item distribution and the assigned item label; a summary of the item content is provided in Table 1.

Data Collection Procedures

Study participants (i.e., students and experienced pharmacists) met with the researcher for a 90-minute one-on-one interview, which consisted of four phases: (1) consent and instructions, (2) think-aloud interview, (3) cognitive interview, and (4) demographic questionnaire. At the beginning of the study—phase one—participants were consented and then instructed on the session procedures, which included audio recording for subsequent data analysis. At the end of the study, participants completed a demographic survey to describe their background and perceptions of empathy. In addition, the demographic survey included the Questionnaire of Cognitive and Affective Empathy (QCAE), which provides a self-reported measure of cognitive and affective empathy.⁴³ Scores on the QCAE can range from 31 to 124; the score is the sum of the cognitive empathy (CE) sub-score (range of 19 to 76) and the affective empathy (AE) sub-score (range of 12 to 48).

In the second phase—the think-aloud interview—participants were instructed to complete the full 12-item SJT one test item at a time; they were not allowed to revisit prior questions once they had finished. The item order was randomized for each participant to minimize order effects. During the think-aloud, participants were instructed to verbalize their thoughts as they worked through the items; the interviewer was only to intervene in the event of silence lasting greater than five seconds and could only use prompts such as “keep talking”.²³ It was important to not alter cognitive processes during this phase; therefore, we did not request individuals to elaborate and describe their approach to protect against introducing biases into participant thought processes.⁴⁴

In the third phase—the cognitive interview—participants were asked about their understanding of and approach to the SJT items. The distinct difference between the think-aloud and cognitive interview is that the latter included questions related to how participants solved each problem and why they made certain

selection decisions. Participants had the opportunity to review each item and their responses as they answered the cognitive interview questions. However, participants were not permitted to change their submitted responses. The interview protocol was structured to address the research questions for the study, which was to learn more about the role of attributes considered to be relevant in the decision-making process during SJTs.

Due to time constraints, each participant was asked about their responses to eight of the twelve items. Items were evenly distributed based on the component of empathy assessed and the setting. In other words, participants completed four items in a healthcare setting, four items in a non-healthcare setting, four items measuring cognitive empathy, and four items measuring affective empathy to varying degrees of overlap. In summary, there were twenty cognitive interviews conducted per item including ten interviews with students and ten interviews with pharmacists.

SJT response data and demographic survey responses were compiled into an electronic database and labeled using the unique participant identifier. Audio files from the interviews were converted to written transcripts using an online transcription service. For the think-aloud interviews, the entire interview was maintained in its presented order and grouped by the level of the participant (i.e., student or pharmacist). For the cognitive interviews, the segments were grouped according to the test item. For example, all cognitive interview questions related to item CH1 were grouped into one transcript for analysis and subdivided based on whether it was a student or a pharmacist to optimize data analyses.

Data Preparation & Analysis Procedures

Models that outline antecedents as well as models that evaluate survey response processes were used to create an initial codebook for coding the transcripts from the cognitive interviews.^{18,26,45-47} The final codebook is provided in the supplemental appendix. The coding process for the cognitive interview included a calibration phase followed by three rounds of coding that were conducted by two researchers. During the calibration phase, a mock transcript was used from the pilot test of four SJT items; the two researchers coded the transcript independently according to the initial codebook and met to review discrepancies, generate example quotes for the codebook, and modify the codebook definitions as needed. The goal of the calibration phase was to allow the raters an opportunity to align coding expectations and resolve concerns prior to the official coding process.⁴⁸

Next, the cognitive interview transcript coding occurred in three rounds using a step-wise approach that involved (1) double-coding by two researchers, (2) auditing by a second researcher after the first researcher completed the coding, and (3) independent coding by only one researcher. This process is a commonly used qualitative strategy for large data sets so that two researchers are not required to code all elements of the data (Saldana, 2016). The step-wise approach allows for frequent calibration and resolution of discrepancies. The coders only progressed to the next round of coding if the rater agreement was above 80%, per common qualitative research standards.^{48,49} The agreement was 80.2% for the first round and 97.7% for the second round, with the coders resolving any discrepancies until 100% agreement

was met. Researchers were also permitted to inductively code segments of text as “other” if they identified what they perceived to be an emerging code that was not identified in the initial codebook. The think-aloud interviews were coded using the same process. During the coding of the think-aloud interviews there were no new codes added to the codebook. Rater agreement was 87.5% agreement during the double-coding round and 94.9% agreement during the auditing round.

The coded transcripts were reviewed to examine the prevalence and context of the participant’s utterances; patterns and relationships among the codes were reviewed. There was evidence to support an underlying SJT response process generated from salient observations in the cognitive and think-aloud interviews. Thus, using these findings and previous models on survey response processes and factors suspected to influence the response process, a new model for the SJT response process was proposed.^{18,26,45,46}

Each SJT item was scored using a partial credit model based on how much the participant’s ranking of the five options differed from the key created by subject matter experts. Each item was worth a maximum of 20 points with a possible 240 points for the total SJT; four points were awarded for each response option that was ranked in the same position designated by the key. The score is reduced for each ranked option based on the distance from the key; this is consistent with the scoring methods for SJTs in the health professions.³⁹

Details about the psychometric qualities of the items and instrument are beyond the scope of this article (a full description is provided elsewhere⁵⁰), but overall the findings demonstrated the SJT provided sufficiently reliable and valid data regarding the assessment of empathy. Demographic data collected from participants were summarized using descriptive statistics and Mann-Whitney tests to compare groups where necessary; the quantitative comparisons are not presented extensively in this paper as the focus was on the exploratory research related to the SJT response process.

Results

Participant Characteristics

The student participants were predominantly female (n = 11, 73.3%) with a median age of 24 years (range 22-45 years). Most students were entering their third or fourth year of pharmacy school (n = 11, 73%), meaning they had experience working in a pharmacy practice setting through required clinical experiences. In addition, 13 students (87%) indicated working in a healthcare-related field outside of their coursework. Eight students (53%) reported working in a non-healthcare human services field with a one year of experience being the median (range 0-10 years). Eighty percent (n = 12) of students reported having training related to empathy; they most often cited coursework or classroom discussions regarding mental health and working with patients.

The pharmacists were predominantly female (n = 13, 86.6%) with a median age of 36 (range 29-51 years). All pharmacists were employed in a university hospital setting, working in various practice areas,

with a median of 8 years of experience as a licensed pharmacist (range 6-23 years). Most completed residency training (n = 13, 87%) and were board certified (n = 11, 73%), indicating these individuals have extensive training in specialty areas and providing advanced patient care. Eleven pharmacists (73%) reported previously working in a non-healthcare human services field with a median of 4 years of experience (range 0-10 years) outside of pharmacy. Only 33% (n = 5) of pharmacists reported having training related to empathy; participants frequently cited exposure to material related to emotional intelligence or service recovery training specific to their institution.

All 30 participants completed the QCAE. Mean QCAE score was 91.8 (SD 6.1) and total scores ranged from 79 to 105. The mean CE and AE sub-scores were 57.1 (SD 5.4) and 34.7 (SD 3.8), respectively. Results of a Mann-Whitney test suggested non-significant differences ($p > .05$) between the median QCAE, CE, and AE scores for participants in the student and pharmacist groups. This finding implies that the pharmacy students and licensed pharmacists included in this study scored similarly and did not differ significantly in their levels of cognitive, affective, and overall empathy.

Average SJT score was 180.6 (SD 11.8) with a range of 142 points (59.2% correct) to 200 points (83.3% correct). Results of a Mann-Whitney test suggested non-significant differences ($p > .05$) between the median SJT scores for participants in the student and pharmacist groups; this corroborates findings from the QCAE that suggested similar student and practitioner empathy levels.

Proposed SJT Response Process Model

In this section, a model has been proposed that builds on the framework of Tourangeau and colleagues, who describe survey response processes as including four key components: comprehension, retrieval, judgment, and response selection.²⁶ This framework was combined with features previously reported to be salient in the response process in addition to new features identified through this exploratory research.^{18,45,46}

The model, provided in Figure 2, includes the four primary components connected to the features that are proposed to influence each step in this process. Features that are bolded are those that have substantial evidence from cognitive and think-aloud interviews to support their existence in SJT response processes (i.e., described in detail in the subsequent sections), whereas those that are not bolded have limited data to support their inclusion. All features evaluated in this exploratory analysis were included as there were references to all components at least once in the process; therefore, the significance of these relationships cannot be excluded. Due to the qualitative nature of this study, a larger sample size would be necessary to confirm if the minor features could be excluded in subsequent models. Within each box connected to the primary component, features are ordered in terms of their prevalence (i.e., features that are higher on the list were referenced more frequently and identified as having a notable influence on the response process).

Comprehension Component

Comprehension was considered an essential component of SJT response processes, as a participant must read the item to be able to answer it accordingly. This component also included references to how participants interpreted key elements of SJT scenarios, which is a significant component of the comprehension process. This research identified two features not previously described in the literature: (1) participants often identified a task or objective that needed to be completed and (2) participants made assumptions about the scenario. In addition, the comprehension component is connected to the ability to identify the construct as the examinee's interpretation of the item can be related to the suspected construct.

First, participants often identified an objective that was to be achieved in the scenario. Provided response options were then evaluated—in the judgment process—based on predictions of how well that response would achieve the targeted objective, among other factors. The objectives identified by participants in the cognitive and think-aloud interviews were categorized based on the goal they described. A list of these categories, descriptions, and examples (i.e., excerpts from the pharmacists and students) are provided in Table 2; these categories are ordered from most to least prevalent.

The task objective most often referenced by participants was related to the exchange of information, which could include collecting or sharing information with another individual. The objective least often described by participants referred to modifying a relationship, often between a patient and the healthcare provider. Of note, many task objectives were broad and lacked a specific focus. For example, participants made general statements about something working well or not without any indication of an explicit goal. Based on the distribution of codes, participants attempted to identify the task objective during this SJT regardless of the item setting, empathy component being assessed, or participant type.

Next, in addition to identifying the objective of SJT scenarios, comprehension of SJT items also included the participant making key assumptions about the presented case. Throughout the cognitive and think-aloud interviews, participants made statements about how they interpreted information that was provided. These assumptions could be classified according to what the assumptions were about, which is summarized in Table 3 with descriptions and examples. The assumption categories are organized from most to least prevalent across all interviews. Based on the distribution of codes, there was some evidence to suggest that the type of component being assessed may contribute to varying uses of assumptions; however, the limited sample size precludes generalization.

In general, assumptions appeared to serve as a component of the response process for some participants when there were insufficient details provided in the scenario. As many of these scenarios were designed to exclude extraneous details, it was possible that this required more inferences by the participants. One participant best described this process by saying, “there’s a fair amount of projection” onto the scenario, depending on the elements that were provided. These data suggest that details about the scenario may be necessary if the use of assumptions in the comprehension process is not desirable. Overall, assumptions made up a small proportion of the total number of codes (3.1%), therefore, there is minimal evidence to suggest that assumptions are an overwhelmingly significant component of the SJT response

process. It is evident, however, that assumptions can be used by participants to fill in the gaps and it may be advisable that SJT design includes explicit statements for examinees pertaining to assumptions about the setting or other features to avoid misinterpretation.

Retrieval Component

Retrieval was the next component of the SJT response process in which participants reflected on knowledge and experiences pertinent to the scenario while they formulated their response selection—this also included references to job-specific and general knowledge and experiences. Of note, in the proposed model, there is a bidirectional relationship between retrieval and judgment that differs somewhat from the original model presented by Tourangeau and colleagues (2000). The proposed model suggests that the response process is not always linear and can integrate various memories and judgments that build on each other prior to the final decision in the response selection, which was evident by participants who retrieved multiple experiences or knowledge elements when discussing SJT items.

Transcripts were analyzed to determine if there were consistent features of the experiences and knowledge referenced by participants that were retrieved during the SJT. References to job-specific and general experiences often included features related to the location, the actors in the scene, and the task or topic. In addition, the experiences could be classified on their similarity to the presented scenario, the specificity of the details provided, and the recency of the memory to the present moment. Features of knowledge references included information, a strategy, or a skill that was applicable to the scenario. Table 4 provides a description of these features and examples from the transcripts.

With respect to job-specific experiences and knowledge, pharmacists and students generally referenced these elements in very similar ways. Pharmacists often explicitly connected to their work experiences with few references to pharmacy school; whereas students included experiences from school, clinical rotations, and some work experiences. In addition, student experiences more often included observations of interactions in which they were not an active participant as well as shared stories, class discussions, and simulations. For example, one student participant (S10) discussed how they had “seen some pharmacists delivering sensitive information about what could happen with certain drugs”; a pharmacist, P13, when discussing the same test item instead thought “about a situation when [they] were practicing in the HIV clinic.” Another example was from S3 who stated, “I know we talked about a lot of different scenarios in class... especially diabetes patients” and S2 who shared, “we’ve talked about medication errors in class a lot and I’ve talked about it on some of my rotations.” The data suggests that students more often integrate job-specific experiences that relate to their education and training witnessed so far, which may not include their direct involvement in a similar scenario.

Components that were not discussed frequently included general experiences and knowledge. In general, the experiences tended to be vague and closely related to the presented SJT scenarios. The actors in these scenarios were often friends and family members and the discussion about these experiences occurred mostly when discussing items referring to non-healthcare settings. One notable feature was that examples from television shows were sometimes referenced as viable experiences. For example, when

P15 was discussing the item related to a friend taking a medication to help them study their immediate response when asked about the question was “Jesse Spano – from Saved by the Bell.” One student, S13, also discussed “I think of experiences that a lot of times I watch on TV shows like Dateline.” General knowledge often included references to information such as “just thinking about social norms, you wouldn’t confront somebody in the grocery store”, as shared by S14. Overall, there is minimal evidence to suggest that general experiences and knowledge include particularly salient features that contributed to SJT response processes in this study.

Participants also included nondescript experiences as well as references to a lack of experience or knowledge; however, these references were limited. Examples included instances where P1 stated “this [question] is a tough one because I feel like this like a reality every day” and S14 who shared “this one felt familiar to me.” References to a lack of experience, however, were reviewed to determine if they were more prevalent in specific scenarios. Most participants made statements of unfamiliarity, such as “I don’t really have very much to draw on” (S3) or “this has never happened” (P14). One difference, however, was that pharmacists tended to be more specific when they considered whether they had experiences to draw from. For example, P6 stated “I haven’t had a particular scenario with regards to chemotherapy” whereas students discussing the same question would state more generally that they “haven’t been in a situation where a family member is that upset” (S3). The data suggest that practitioners may be more attentive to granular details compared to students when searching for similar experiences.

Judgments Component

Judgments represented the most prominent code and included comments about the decision-making process as well as any value statement made while assessing the response options. The analysis for this component was focused on factors of SJT frameworks that pertained to the judgments, such as emotional intelligence, self-perception, ability, and impressions management. In addition, three new sub-codes were identified during the analysis: perceptions, feelings about the test, and context.

One of the most prominent judgments included the use of emotional intelligence, which was defined as the capacity to be aware of, control, and express one’s emotions as well as the emotions of others (Cherry, Fletcher, O’Sullivan, & Dornan, 2014). This was not abnormal as this SJT was intended to measure participant empathy. Explicit references to affective and cognitive empathy separately, however, were relatively infrequent across interviews compared to other codes. Instead, broad references to empathy were exhibited more often. The remaining factors in SJT frameworks—self-perception, impressions management, and ability—were infrequently discussed among the interviews but were still included in the model as they pertained to judgments in SJT response processes and were consistent with theoretical frameworks about SJTs. Of these three codes, self-perception was the most common, whereas impression management and ability were rare.

Self-perceptions shared by participants often focused on either: (1) attributes of their personality, (2) their identity as a healthcare provider, friend, or family member, or (3) their comfort with a presented scenario. References to their participant personality often included comments such as, “I think I’m probably a little

bit less aggressive” (P11) or “I’m not very confrontational” (S11). References to participant identity typically related to their status as a healthcare provider, such as P07 who stated, “I guess being a pharmacist though, it’s a little clearer”. These references also included their identities outside of work as well, such as when P03 shared that “as a new parent” there are differences in how they perceived some situations. Lastly, some participants were aware of their comfort with engaging in certain scenario. For example, S02 stated “I’d feel more comfortable talking about the error if it was something like food”. Each of these types of self-perceptions contributed to their judgements about the scenario and could impact their response selection; however, overall there was limited evidence to suggest their criticality in the process.

Impression management and ability were less frequent in the response process—most participants reported they forgot that they were told to imagine that the test was being used for selection into a health professions program or residency. For the participants who did not forget, they described a struggle with differentiating their answer choices on what they should do compared to what they would do as expected by the individual administering the test. For example, S12 shared they “kind of knew what the right answer was versus what [they] would actually do was harder to separate.” With regard to ability, participants most often made references to a lack of a knowledge of skill set that would allow them to operate best in the given scenario instead of affirmations about their abilities to succeed in a situation. For example, P07 stated that, “as a pharmacist, I’m not really trained to walk-through the risks and benefits in that case.” Overall, the few references to abilities limited the analysis; however, the factor was still retained within the model as there was some evidence to suggest ability (or the lack thereof) may be play a role in the response process in that some response options were ranked lower if the participant did not feel they had the skill set necessary to successfully carry out a response option.

Another new feature identified inductively was that participants made references to perceptions of factors weighed when evaluating response options. These perceptions were coded throughout the cognitive and think-aloud interviews, then categorized based on the features that were most salient. Table 5 includes a summary of the most prevalent categories, as well as a description and example for reference.

The most prevalent type of comment from participants was regarding the impact on how others would think about them if a certain response option was selected. Participants most frequently identified negative attributes about the impact on their image including thoughts that it could: “make you look like a jerk” (S10), “come off like accusing the patient” (S03), and “seem unprofessional” (P06). In general, there was a significant concern about how nice a response was or perceptions about the tone in which something was delivered, which could subsequently impact their image and response selection. Examples included comments about response options that “sounded really cold” (S15) or that “can come off a little harsh” (P05); these responses were then not as highly ranked. Similar to this was the perceived integrity of certain response options; for example, participants evaluated if the response was an honest reflection of the situation or if the response was legal. Each could potentially have implications for the image, but these focused specifically on an important element other than how professional or how nice they were coming across. Other perceptions included an awareness of what individuals would do in real-

life scenarios, as well as a balance between perceptions of what participants believed individuals would want in the scenario along with what they would want in the scenario. Lastly, some individuals referenced their instincts in the scenarios and stated, “it just feels right” as their reasoning.

In addition, the setting and contextual features of the item appeared to have an effect on the response process. Interestingly, there were many uses of the phrase “it depends” (and other equivalents) by participants across the transcripts, which suggested the importance of contextual elements in SJT response processes. These factors were coded and classified into four categories: (1) factors pertaining to the participant or *examinee*, (2) factors pertaining to *actors* in the presented scenario, (3) factors pertaining to the *relationship* between the examinee and actors, and (4) factors pertaining to the *situation*.

Participants often cited multiple factors that influence their response process and that these factors could affect their response differently based on the scenario. For example, item AH1 asks how the participant should respond to a patient who is upset about the recent loss of a loved one. One pharmacist, P06, stated that “If it were a friend, I would have been more inclined to share my own personal experiences...I’d feel more comfortable sharing personal loss and talking about it on a more personal level.” In this case, the participant identified that the actor (e.g., a friend instead of a patient) has an impact on the response selection as well as the relationship (e.g., a personal instead of a professional relationship).

Participants commonly identified that relationships with friends and family members come with different expectations compared to relationships with work colleagues or patients. For example, student S10 shared that when trying to convince a patient about not taking a non-prescribed medication compared to convincing a friend, they thought “it’d be easier because you could come at it from the standpoint of I’ve had training in this... and there’s no evidence to back this up or that’s illegal.” In this case, factors such as the examinee’s training as well as the legality of the situation also contribute to the response process. Altering the question to exclude factors such as the illegality or the examinee’s position could alter their response. The impact of the setting on selection of responses was part of another research question of the larger study and is reported in greater detail elsewhere (Wolcott, 2018).

Response selection

The last component of SJT response processes is the response selection, which, for this study, included having the examinees rank their response option on the SJT. Response selection in this study included any reference to the final ranking assigned to any response option. A notable feature of the response selection in this research study was the integration of general strategies that participants reported using throughout the SJT.

Table 6 summarizes the different strategies that were used by participants in making their final selections. In general, most participants approached the response process in the way they were instructed to, which was to rank responses from most to least appropriate. Others, however, considered working

backwards in some situations or identifying the extremes (most and least appropriate) first and then filling in the remaining ranks. Other strategies included comparing response options, guessing, and using a process of elimination. Some participants when reading questions aloud also rephrased the item by orienting themselves within the question. One pharmacist, for example, started each response option with "Do you..." when reading the item aloud despite this not being present in the written document. Students appeared to reference test taking strategies more often than pharmacists based on the distribution of codes. In summary, there was some evidence to suggest that general test taking strategies are a relevant feature in SJT response processes and the use of strategies may differ based on who is taking the test.

Discussion

The use of SJTs in the health professions is a rapidly growing phenomenon as an approach to measure professional competence.² Prior research about survey response processes suggested that the cognitive process during an SJT would be similar. This was suspected to include elements of comprehension, retrieval, judgment, and response selection.^{18,26} Results from this study provided evidence that these four components are indeed present in SJT response processes according to utterances in the cognitive and think-aloud interviews. Statements related to judgments, retrieval of memories or information, and response selections were some of the most prevalent codes in the qualitative analysis of the data. In addition, these features can independently contribute to the decision-making process and can therefore influence score interpretations if any of these features are inappropriately influenced.

The findings highlight the important factors that contribute to response processes during an SJT. Building on previous SJT research that has identified multiple antecedents suspected to influence response selections (e.g., job-specific and general experiences;), our framework included individual characteristics (e.g., emotional intelligence, ability, personality, etc.) that influence the decision-making process.⁴⁶ Results from this study suggest there are multiple factors that are considered by participants during the response process, which can vary greatly among examinees in the extent to which they are applied. Specifically, the findings confirmed that job-specific experiences and knowledge as well as emotional intelligence were salient features of SJT response processes. Other features such as general experience and knowledge, self-perceptions, ability, and impressions management were not sufficiently supported as pertinent components of the response process. The proposed model included all of these features, though, as the lack of utterances about a particular feature was not considered to be sufficient evidence to discard it in this exploratory phase. Additional research is necessary to confirm the findings of this research.

Results pertaining to this study also identified new features of the response process that have not been previously described in the literature. Specifically, results suggested that participants often attempt to identify a task objective during this SJT and evaluate how well response options achieve that task based on their comprehension of the elements that are presented. In addition, they often make assumptions about the scenario that influence how they comprehend the situation. During the judgment process, participants identified that they evaluated response options according to their perceptions on how the

action would reflect on their image, whether it was something they could imagine doing in real life, or what they would want done for them in the situation. Moreover, participants identified that contextual features such as the item setting could greatly influence their response selections. Lastly, there were a host of test-taking strategies that participants employed during this SJT that may be broadly applicable regardless of the item.

In general, these new features have not been extensively discussed in prior research about SJT response processes. For example, Rockstuhl and colleagues were the first to report evidence about SJT response processes; however, they categorized participant utterances simply on the content presented.¹⁹ This research extends on this prior work by addressing how these features relate to the four-component model of the response process that was evident and describing these features in greater detail. In addition, Krumm and colleagues presented a small research study that identified some of the strategies test-takers used when completing an SJT.²⁰ Our study took that research further by identifying additional strategies and describing how participants evaluated the effectiveness of response options better. The previous work in this area was limited in the depth of information it provided about SJT response processes. The results of this research question, therefore, greatly expanded the overall understanding of these features within the response process.

Implications and Limitations

This is the first study to explicitly show which components of the response process were salient using a novel integration of cognitive and think-aloud interviews, which allows for the assessment of comprehension and retrieval processes in a step-wise approach. In addition, the results were used to generate a model that can be tested through future research and be used as a mechanism to generate validity evidence for SJTs. The study takes the first step in generating critical validity evidence about the response process that has been substantially lacking in SJT research. In addition, the study presents a comprehensive, in-depth analysis of the response process using qualitative strategies that offers unique content and insight into the phenomena, which has not been extensively used previously.

A limitation of the presented work is that—in this model—the relationship between the individual components is not fully specified as the focus of this research was to explore the response process holistically. Future research, however, could investigate which components are most influential in SJT performance, how they relate to one another as well as other variables, and whether they influence multiple components of the response process instead of the single component structure provided here. Furthermore, this model has been constructed using an SJT intended to measure empathy. As such, this model may not be broadly applicable to other constructs evaluated using an SJT. It is, however, the first step in developing a more comprehensive and integrated model than previously documented in the literature, which could be tested with SJTs on other topics. Furthermore, the study is limited as it focuses on response processes articulated by participants from one profession (i.e., pharmacy) and one region (i.e., southeastern United States). Due to the exploratory nature of this study, additional research is

necessary to identify if the model can be applied in other health professions settings or regions with differing experiences or practices.

Conclusion

SJT response processes include four key components: comprehension, retrieval, judgment, and response selection. This is the first research that has explicitly shown these components are present and to offer a model that integrates these features from the literature with evidence. Moreover, this research identified factors that contribute to each of these components in the response process. Five new features not included in previous research were identified in this study: (1) identification of the task objective, (2) assumptions about SJT scenarios, (3) perceptions of the response options, (4) contextual features of the response options, and (5) general strategies in the response selection. This study provides a comprehensive evaluation of the SJT response process and contributes to foundational attributes of the validity evidence of SJTs. These findings can inform the design, interpretation, and utility of SJTs.

Declarations

This study was reviewed and approved by the University of North Carolina Institutional Review Board. All participants consented to their engagement in the study.

The datasets generated and/or analyzed during the current study are not publicly available but are available from the corresponding author on reasonable request.

The authors have no financial or non-financial competing interests to disclose.

This study was not funded by an external or internal entity.

MW was responsible for study design, implementation, data analysis, and creation of the manuscript. NL was responsible for data analysis and review of the manuscript. JZ was responsible for data analysis and review of the manuscript. JM was responsible for reviewing study design and review of the manuscript.

References

1. Koczwara A, Patterson F, Zibarras L, Kerrin M, Irish B, Wilkinson M. Evaluating cognitive ability, knowledge tests, and situational judgment tests for postgraduate selection. *Med Educ.* 2012;46:399-408.
2. Patterson F, Knight A, Dowell J, Nicholson S, Cousans F, Cleland J. How effective are selection methods in medical education? A systematic review. *Med Educ.* 2016;50:36-60.
3. Champion MC, Ployhart RE, MacKenzie Jr WI. The state of research on situational judgment tests: A content analysis and directions for future research. *Human Performance.* 2014;27:283-310.

4. Chan D, Schmitt N. Situational judgment and job performance. *Human Performance*. 2002;15:233-254.
5. Lievens F, Patterson F. The validity and incremental validity of knowledge tests, low-fidelity simulations, and high-fidelity simulations for predicting job performance in advanced-level high-stakes selection. *J Appl Psychol*. 2011;96:927-940.
6. De Leng WE, Stegers-Jager KM, Husbands A, Dowell JS, Born MP, Themmen APN. Scoring methods of a situational judgment test: Influence on internal consistency reliability, adverse impact and correlation with personality? *Adv Health Sci Educ Theory Pract*. 2017;22:243-265.
7. Colbert-Getz JM, Pippitt K, Chan B. Developing a situational judgment test blueprint for assessing the non-cognitive skills of applicants at the University of Utah School of Medicine the United States. *J Educ Eval Health Prof*. 2015;12:51-55.
8. Patterson F, Ashworth V, Zibarras L, Coan P, Kerrin M, O'Neill P. (2012) Evaluations of situational judgement tests to assess non-academic attributes in selection. *Med Educ*. 2012;46:850-868.
9. Petty-Saphon K, Walker KA, Patterson F, Ashworth V, Edwards H. Situational judgment tests reliably measure professional attributes important for clinical practice. *Adv Med Educ Pract*. 2016;8: 21-23.
10. Wolcott MD, Lupton-Smith C, Cox WC, McLaughlin JE. A 5-minute situational judgment test to assess empathy in first year student pharmacists. *Am J Pharm Educ*. 2019:doi.org/10.5688/ajpe6960.
11. McDaniel MA, List SK, Kepes S. The "hot mess" of situational judgment test construct validity and other issues. *Industrial and Organizational Psychology*. 2016;31(1):47-51.
12. Sorrel MA, Olea J, Abad FJ, de la Torre J, Aguado D, Lievens F. Validity and reliability of situational judgment test scores: A new approach based on cognitive diagnosis models. *Organizational Research Methods*. 2016;19(3):506-532.
13. Caines J, Bridglall BL, Chatterji M. Understanding validity and fairness issues in high-stakes individual testing situations. *Quality Assurance in Education*. 2014;22(1):5-18.
14. American Educational Research Association, American Psychological Association, National Council on Measurement in Education. *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association; 2014.
15. Fan J, Stuhlman M, Chen L, Weng Q. Both general domain knowledge and situation assessment are needed to better understand how SJTs work. *Industrial and Organizational Psychology*. 2016;31(1):43-47.
16. Harris AM, Siedor LE, Fan Y, Listyg B, Carter NT. In defense of the situation: An interactionist explanation for performance on situational judgment tests. *Industrial and Organizational Psychology*. 2016;31(1):23-28.
17. Melchers KG, Kleinmann M. Why situational judgment is a missing component in the theory of SJTs. *Industrial and Organizational Psychology*. 2016;31(1):29-34.
18. Ployhart RE. The predictor response process model. In: Weekly JA, Ployhart RE, editors. *Situational judgement tests: Theory, measurement, and application*. Mahwah, NJ: Lawrence Erlbaum; 2006. p. 83-105.

19. Rockstuhl T, Ang S, Ng KY, Lievens F, Van Dyne L. Putting judging situations into situational judgment tests: Evidence from intercultural multimedia SJTs. *J Appl Psychol.* 2015;100:464-480.
20. Krumm S, Lievens F, Huffmeier J, Lipnevich AA, Bendels H, Hertel G. How “situational” is judgment in situational judgment tests? *J Appl Psychol.* 2015;100:399-416.
21. Pellegrino J, Chudowsky N, Glaser R. Knowing what students know: The science and design of educational assessment. Board on Testing and Assessment, Center for Education, National Research Council, Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academic Press; 2001.
22. Nichols P, Huff K. Assessments of complex thinking. In: Ercikan K, Pellegrino JW, editors. Validation of score meaning for the next generation of assessments: The use of response processes. New York, NY: Routledge; 2017. p.63-74.
23. Leighton JP. Using think-aloud interviews and cognitive labs in educational research: Understanding qualitative research. New York, NY: Oxford University Press; 2017.
24. Willis GB. Analysis of the cognitive interview in questionnaire design: Understanding qualitative research. New York, NY: Oxford University Press; 2015.
25. Schwarz N. Cognitive aspects of survey methodology. *Applied Cognitive Psychology.* 2007;21:277-287.
26. Tourangeau R, Rips LC, Rasinski K. The psychology of survey response. Cambridge, MA: Cambridge University Press; 2000.
27. Brooks ME, Highhouse S. Can good judgment be measured? In: Weekly JA, Ployhart RE, editors. Situational judgement tests: Theory, measurement, and application. Mahwah, NJ: Lawrence Erlbaum; 2006. p. 39-56.
28. Kim SS, Kaplowitz S, Johnston MV. The effects of physician empathy on patient satisfaction and compliance. *Eval Health Prof.* 2004;27(3):237-251.
29. Riess H. The science of empathy. *J Patient Exp.* 2017;4(2):74-77.
30. Decety J, Jackson PI. The functional architecture of human empathy. *Behavioral and Cognitive Neuroscience Reviews.* 2004;3(2):71-100.
31. Hojat M. Empathy in patient care: Antecedents, development, measurement, and outcomes. New York, NY: Springer; 2007.
32. Quince T, Thiemann P, Benson J, Hyde S. Undergraduate medical students’ empathy: Current perspectives. *Adv Med Educ Pract.* 2016;7:443-455.
33. Tamayo CA, Rizkalla MN, Henderson KK. Cognitive, behavioral, and emotional empathy in pharmacy students: Targeting programs for curriculum modification. *Front Pharmacol.* 2015;7:Article 96.
34. Fjortoft N, Van Winkle LJ, Hojat M. Measuring empathy in pharmacy students. *Am J Pharm Educ.* 2011;75:Article 109.
35. Nunes P, Williams S, Sa B, Stevenson K. A study of empathy decline in students from five health disciplines during their first year of training. *Int J Med Educ.* 2011;2:12-17.

36. Lievens F. Construct-driven SJTs: Toward and agenda for future research. *International Journal of Testing*. 2017;17(3):269-276.
37. McDaniel MA, Hartman NS, Whetzel D, Grubb III WL. Situational judgment tests, response instructions, and validity: A meta-analysis. *Personnel Psychology*. 2007;60(1):63-91.
38. McDaniel MA, Nguyen NT. Situational judgment tests: A review of practice and constructs assessed. *International Journal of Selection and Assessment*. 2001;9(1):103-113.
39. Patterson F, Zibarras L, Ashworth V. Situational judgement tests in medical education and training: Research, theory, and practice: AMEE Guide No. 100. *Med Teach*. 2016;38(1):3-17.
40. Weekley JA, Ployhart RE. An introduction to situational judgment testing. In: Weekly JA, Ployhart, RE, editors. *Situational judgement tests: Theory, measurement, and application*. Mahwah, NJ: Lawrence Erlbaum; 2006. p. 1-10.
41. Patterson F, Baron H, Carr V, Plint S, Lane P. Evaluation of three short-listing methodologies for selection into post-graduate training: The case of General Practice in the UK. *Med Educ*. 2009;43:50-57.
42. Siegel S, Castellan NJ. *Nonparametric statistics for the behavioral sciences*, New York, NY: McGraw-Hill; 1988.
43. Renate LE, Corcoran R, Drake R, Shryane NM, Vollm BA. The QCAE: A questionnaire of cognitive and affective empathy. *Journal of Personality Assessment*. 2011;93(1):84-95.
44. Chi MTH. Quantifying qualitative analyses of verbal data: A practical guide. *The Journal of Learning Sciences*. 1997;6:271-315.
45. Griffin B. The ability to identify criteria: Its relationship with social understanding, preparation, and impression management in affecting predictor performance in a high-stakes selection context. *Human Performance*. 2014;27:147-164.
46. Lievens F, Motowidlo SJ. Situational judgment tests: From measures of situational judgment to measures of general domain knowledge. *Industrial and Organizational Psychology*. 2016;9(1):3-22.
47. Chessa AG, Holleman BC. Answering attitudinal questions: Modelling the response process underlying contrastive questions. *Applied Cognitive Psychology*. 2007;21:203-225.
48. Saldana J. *The coding manual for qualitative researchers*. Thousand Oaks, CA: SAGE; 2016.
49. Merriam S, Tisdell EJ. *Qualitative research: A guide to design and implementation*. San Francisco: Jossey-Bass; 2016.
50. Wolcott MD. The situational judgment test validity void: Describing participant response processes [doctoral dissertation]. University of North Carolina; 2019 Retrieved from ProQuest Dissertations & Theses., Accession Number 10981238.

Tables

Due to technical limitations, all tables are only available for download from the Supplementary Files section.

Figures

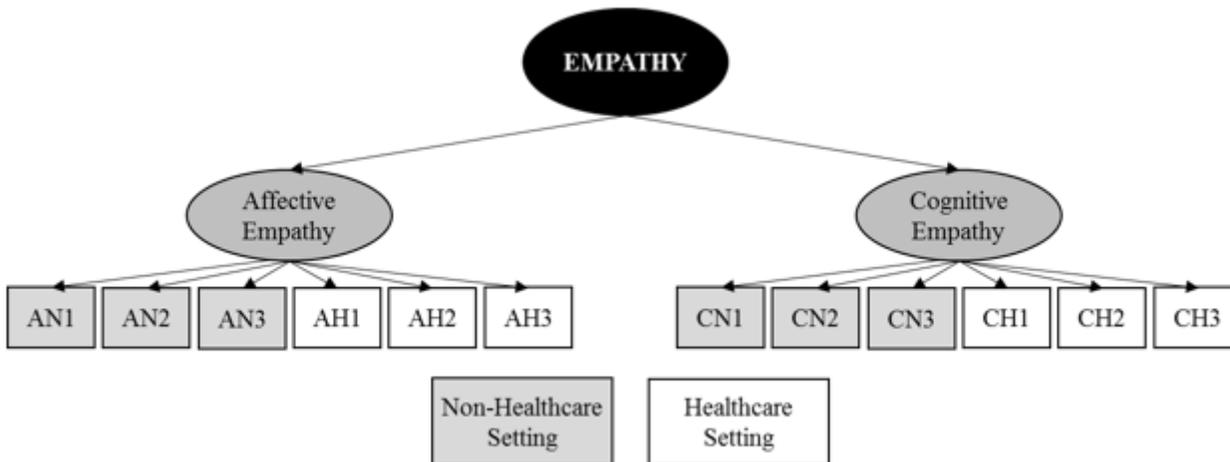


Figure 1

Map of SJT items, settings, and the associated construct components. A = Affective Empathy; C = Cognitive Empathy; H = Healthcare Setting, N = Non-Healthcare Setting; 1, 2, 3 = Item Number

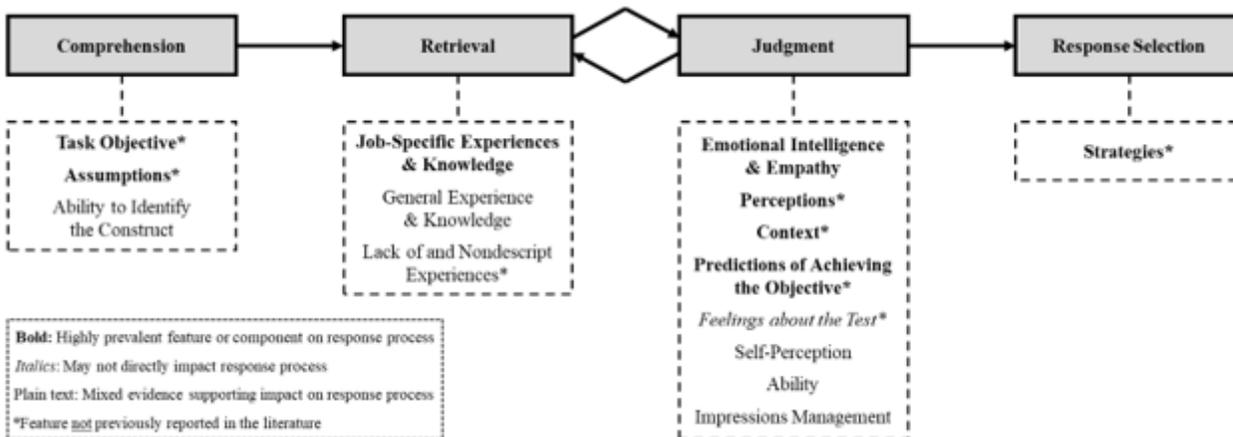


Figure 2

Proposed model of SJT response processes

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementalMaterial.docx](#)
- [Tables.docx](#)