Using natural language processing to measure cognitive load during use-of-force decision-making training

Vivian Ta-Johnson
Department of Psychology, Lake Forest College, Lake Forest, Illinois, USA
Joel Suss
Department of Psychology, Wichita State University, Wichita, Kansas, USA, and
Brian Lande
Polis Solutions, Inc., Dallas, Texas, USA

Received 12 June 2022
Revised 15 November 2022
28 November 2022
Accepted 30 November 2022

Abstract

Purpose – Few studies have tested the efficacy of instruction based on cognitive load theory in police use-of-force (UoF) training due to limitations of existing cognitive load measures. Although linguistic measures of cognitive load address these limitations, they have yet to be applied to police UoF training. This study aims to discuss the aforementioned issue.

Design/methodology/approach – Officers' verbal behavioral data from two UoF de-escalation projects were used to calculate cognitive load and assess how it varied with officer experience level (less-experienced, experienced). The verbal data were further analyzed to examine specific thinking patterns that contributed to heightened cognitive load across officer experience levels.

Findings – Across both studies, responses from less-experienced officers contained greater usage of cognitive language than responses from experienced officers. Specific cognitive processes that contribute to cognitive load in specific situations were also identified.

Originality/value – This paper enables police trainers to facilitate the development of adaptive training strategies to improve police UoF training via the reduction of cognitive load, and also contributes to the collective understanding of how less-experienced and experienced officers differ in their UoF decision-making.

Keywords Language, Decision making, Experience, Use of force, Police training

Paper type Research paper

Use-of-force (UoF) training has been an integral component of police training programs for more than a century (Arnspiger and Bowers, 1996). Much attention has been placed on identifying new approaches to improve UoF training given that incidents involving inappropriate UoF by police continue to occur and have serious consequences on police–community relations (Ross, 2000). One such approach incorporates cognitive load theory (CLT), a validated theory of instructional design focusing on designing/delivering instructional materials that produce optimal learning conditions for learners (Mugford *et al.*, 2013). According to CLT, optimal learning occurs when learners experience appropriate levels of cognitive load – the amount of working-memory demand induced in a particular task (Sweller, 1988) – during training. Too much cognitive load can overwhelm the learner's processing capacity, hinder the learning process and eliminate beneficial effects of training (Van Merrienboer and Sweller, 2005). Experience level affects the degree of cognitive load that

This work was supported by the Bureau of Justice Assistance (BJA-2016-VI-BX-K005). The authors acknowledge Paul Ward's guidance in the design of the study that generated the data analyzed in Study 2.

Policing: An International Journal

1363-951X
DOI 10.1108/PIJPSM-06-2022-0084

learners experience during training, making it necessary to account for learners' experience level when designing/delivering instructional materials (Ayres and Paas, 2009). As such, there have been calls to incorporate CLT-based instruction into UoF training to reduce cognitive load and maximize learning outcomes (Bennell *et al.*, 2007).

However, few studies have tested the efficacy of CLT-based instruction in UoF training (Mugford *et al.*, 2013). This might be due to the limitations of existing measures of cognitive load (Beckmann, 2010). Obtaining real-time, non-invasive measures of cognitive load is a challenging task. Existing cognitive load measures pose methodological issues (lack of standardization, need for metacognitive insight, high cost, steep learning curve, restriction of body movement and replication challenges) that can interfere with instructional procedures and interrupt the learning process (Khawaja *et al.*, 2014). Linguistic measures of cognitive load—measures of cognitive load based on verbal behavioral data—address these shortcomings and provide methodological advantages that are not available in other cognitive load measures (Khawaja *et al.*, 2014). Verbal behavioral data involves the use of words to transmit information and encompasses both spoken and written data (Lemke, 2012). Linguistic measures of cognitive load have regularly been used in other disciplines such as human factors, human–computer interaction and medicine (Konopasky *et al.*, 2020), but have yet to be applied in police training. We addressed this limitation in the current paper by using officer verbal behavioral data from two separate UoF de-escalation projects to calculate cognitive load and assessed whether cognitive load varied across officer experience level (less-experienced, experienced). Additionally, the verbal data were further analyzed to gain insight into specific thinking patterns that contributed to heightened cognitive load across officer experience levels in each UoF de-escalation project.

This paper makes important practical and research contributions to police training. First, we demonstrate how a validated linguistic approach to measuring cognitive load can provide robust, non-intrusive assessments and inform the cognitive processes underlying officer cognitive load in UoF training. This provides police trainers with the ability to more easily manage trainees' performance during training, and, as a result, directly monitor how well trainees are understanding the content and adjust training interventions to maximize content retention as needed. Second, assessing cognitive load across officer experience levels provides insight into the differences by which less-experienced and experienced officers mentally process situations involving UoF. Compared to less-experienced individuals, experienced individuals tend to make decisions that are more efficient, effective and accurate (Klein *et al.*, 2017). However, research directly comparing UoF decision-making among less-experienced and experienced officers has been limited. More research is needed to better understand all the components that make expert performance in UoF decision-making "expert", which is marked by the ability to use one's skill set to navigate complex, novel situations and make effective, accurate and efficient decisions (Suss and Boulton, 2019). Moreover, the ability to identify the specific thinking patterns that contribute to heightened cognitive load during training enables police trainers to target particular mental processes that help or hinder the learning process. Overall, this provides police trainers with the tools to improve the training process and enhance the transfer of skills to real-world situations, thereby facilitating a more adaptive approach to improving police UoF training through cognitive load reduction.

This paper begins with a brief review of the cognitive load literature and the effects of experience level on cognitive load in the context of training. Then, we summarize the common ways in which cognitive load has been assessed and the advantages of using linguistic measures of cognitive load. The goals, methods and results of two UoF de-escalation projects are reported in which officer verbal data are analyzed to assess cognitive load and inform the specific thinking patterns that contribute to cognitive load among less-experienced and experienced officers. Last, the practical and research implications on UoF training for police practitioners and future research are discussed.

Background literature

Cognitive load refers to the amount of working-memory demand induced on an individual in a particular task (Sweller, 1988). Complex tasks that present a substantial amount of novel information put high demands on working memory, increase cognitive load, and thus require more processing capacity and effort to complete. Because working memory can only contain a limited amount of information for a limited amount of time, high cognitive load hinders the ability to process novel information. Performing tasks under high load has negative effects on task completion and overall performance (Sweller *et al.*, 2011). High load also interferes with the learning process (Mugford *et al.*, 2013). As such, "in order for . . . training to be effective, instructional methods must facilitate the acquisition and automation of task-relevant schemas without overwhelming the limited processing capacity of the learner" (Bennell *et al.*, 2007, p. 35). In other words, training techniques must be appropriately structured and attuned to each learner's cognitive capacities and schema development for effective learning (Ayres and Paas, 2009). If not, training techniques can lose their effectiveness and jeopardize the entire learning process (Paas *et al.*, 2003). Having the capacity to comprehensively process the training material is also required to transfer newly-learned skills from working memory into long-term memory (Bennell *et al.*, 2020).

A learner's experience level influences the amount of cognitive load they experience in a learning environment. For a given task, individuals with more domain experience will typically experience less cognitive load compared to individuals with less domain experience. This is because the acquisition and automation of schemas—efficient "cognitive structures that serve to organize information into meaningful concepts" (Tennyson and Volk, 2015, p. 709)—have already occurred in police officers with more domain experience. However, cognitive load will decrease as less-experienced officers develop relevant schemas and gain more domain experience on the job (Moreno, 2004). Officers with more experience can identify and process cues to obtain a deep understanding of situations and make more plausible inferences at a faster rate given that they experience less load during decision-making compared to less-experienced officers (Persky and Robinson, 2017). Greater experience facilitates the development of expertise in law enforcement which is characterized by ". . . the ability to adaptively apply one's skills, knowledge, and attributes to novel and complex (e.g. uncertain, time-pressured, dangerous) situations and environments" (Suss and Boulton, 2019). This includes assessing situations accurately (Suss and Ward, 2015), generating quality potential responses (Suss and Ward, 2018) and predicting likely situational outcomes (Suss and Ward, 2015, 2018)—qualities that facilitate apt and appropriately timed courses of action (Ward and Williams, 2003).

Cognitive load measurement

Cognitive load can be measured using a variety of methods, including subjective (Paas *et al.*, 2003), physiological (Kennedy and Scholey, 2000) and performance-based (Paas *et al.*, 2008) methods. Subjective methods commonly take the form of self-report measures in which the respondent rates their own cognitive load. Although self-report measures are a relatively easy and non-obtrusive way to measure cognitive load, the lack of a standardized format raises validity issues (De Jong, 2010). Moreover, respondents may not have the metacognitive insight to accurately assess the amount of cognitive load they experienced (Joseph, 2013). Subjective methods also include third-party judgments of cognitive load. For example, subject-matter experts may observe an individual completing a scenario and rate the amount of cognitive load exhibited by that individual (Khawaja *et al.*, 2009). Physiological measures require the use of equipment such as eye-tracking devices or EEGs. These are impractical for police training, given that such devices are often expensive, restrict natural body movement and require a steep learning curve to operate correctly (Amadieu *et al.*, 2009). Performance-based measures are task-specific and vary across and within tasks, making it difficult to replicate cognitive load measurements across different tasks (Dan and Reiner, 2017).

Linguistic measures of cognitive load are advantageous because they can typically be collected without interfering with training procedures or interrupting the subject's attention, thus providing safeguards that would otherwise introduce noise into the data (Khawaja *et al.*, 2014). For example, learners' spoken/written verbal behaviors during training can be captured and analyzed to calculate cognitive load. These measures are derived directly from learners' behaviors rather than from self-report measures or subjective third-party judgments. Language and verbal behavior are reflective of people's ongoing psychological states and many studies have used verbal behaviors to assess cognitive states (Tausczik and Pennebaker, 2010). Recent technological advancements have also made extremely accurate linguistic models of verbal behavior more accessible to researchers, thereby making it easier than ever before to examine the cognitive processes that were once extremely difficult to capture—including cognitive load.

The current paper

Officer verbal behavioral data from two UoF de-escalation studies were used to assess cognitive load and inform the specific thinking patterns that contributed to heightened cognitive load among less-experienced officers (who did not yet have any operational police other than their recruit training) and experienced officers (who had more than five years of experience as a police officer). Data used in Study 1 were from a UoF de-escalation training project funded by the Bureau of Justice Assistance. A sample of less-experienced and experienced officers [1] completed five decision-making exercises in which they observed police body-worn-camera (BWC) footage and then described (in textual form) the course of action they would take, the information they paid attention to and their assessment of the situation. Due to logistical issues involving the participating law enforcement agencies, we were not able to recruit less-experienced and experienced officers from the same department. Hence this study was primarily used for methodological innovation rather than for seeking to explain group differences. We elaborate more on this in the Method section; in the Discussion section we describe new research we have underway to rectify the sampling issues we encountered and extend the method developed in Study 1 to Study 2. Data used in Study 2 were from a UoF situation-assessment and response-selection study (Suss and Ward, 2018) in which a sample of less-experienced and experienced officers observed video scenarios of the type used in police judgment-and-decision-making simulators. After each scenario, officers were guided through a video-stimulated recall procedure in which they described (in spoken form) the information they paid attention to, how they evaluated the information, the course of action they would take, what they anticipated would happen next and their prior knowledge/experience that influenced their thinking.

Linguistic behavior changes when individuals perform tasks under high versus low cognitive load (Khawaja *et al.*, 2014). Specifically, greater cognitive language usage indicates greater engagement in active cognitive processing such as thinking and reasoning. As such, differences in cognitive load between less-experienced and experienced officers should be reflected in responses after observing footage. People under greater cognitive load experience greater mental effort compared to those under less cognitive load (Sweller *et al.*, 2011). Therefore, responses from less-experienced officers (who experience greater cognitive load) should contain a greater degree of cognitive language than responses from experienced officers (who experience less cognitive load).

Finally, cognitive processing encompasses several internal mental tasks (Smith and Kelly, 2015), and although measures of cognitive load capture the amount of cognitive effort—the degree to which an individual is cognitively engaged with a task (Westbrook and Braver, 2015)—

that is induced, they do not indicate the type(s) of cognitive process(es) the individual is engaging in that is contributing to their cognitive load. For instance, is it the individual's attempt at explaining the factors behind an event (causation) that is contributing to their cognitive load, or is it their attempt at differentiating between cues (differentiation)? We addressed this by analyzing responses from officers across six sub-measures of cognitive language that represent engagement in a psychologically-meaningful cognitive process (Table 1). The specific cognitive processes that underlie cognitive load were not expected to be identical in both studies given that study materials, procedures and scenarios differed across both studies. Instead, an exploratory approach was taken with results informing how specific thinking patterns contribute to cognitive load within each study, and how they differ between less-experienced and experienced officers across the two studies.

Study 1

Method

Participants. Participants were 36 less-experienced officers and 42 experienced officers from two urban police departments in the USA. Less-experienced officers, who were completing their basic recruit training, were recruited from a mid-sized police department. They did not yet have any operational experience working as police officers. Experienced officers were recruited from a large police department and identified as UoF experts by their department's commanders based on criteria that distinguished them as experts (experience, reputation, training, achievement of specialty assignment and being UoF/firearms/defensive tactics instructors). We initially intended to sample less-experienced officers from this large department to control for cultural, organizational and training differences that would confound our findings. Due to operational tempo, the department was unable to provide access to less-experienced officers. This limitation will be addressed in the Discussion.

Procedure of force de-escalation training project. Study materials were presented using Qualtrics. Officers accessed the survey on desktop computers or laptops; officers who completed the training in a group setting were provided with headphones. Officers first provided demographic information and then observed five police incidents captured by BWCs. Each incident depicted a different police–citizen encounter across the USA; videos were obtained from YouTube. Video footage of each incident was divided into three consecutive segments. At the end of each segment, officers were prompted to answer three questions about the segment they just watched by typing their responses in text boxes. These questions are often used to evaluate performance in law enforcement training and captured the course of action the officer would take (rapid decision), the kinds of information the officer paid attention to (critical cues) and the officer's evaluation of the situation (assessment) [2].

Officers also answered additional questions about the footage that were not pertinent to the goals of our current investigation (see https://osf.io/wujkz/for full questionnaires and videos). Officers completed all five exercises in 135 min or less and did not report fatigue.

Sub-category	Greater usage reflects the speaker's . . .	Example terms
Causation	. . . attempts to explain something that gives rise to an effect	*because, effect*
Insight	. . . increased awareness or deeper understanding of something	*think, know*
Tentativeness	. . . uncertainty about something	*maybe, perhaps*
Certainty	. . . increased assurance about something	*always, never*
Discrepancy	. . . contrasting across two or more entities	*should, would, could*
Differentiation	. . . distinguishment between entities, people, or ideas	*but, without, hasn't*

Table 1. Sub-measures of cognitive language

Procedure of the current investigation. Officer responses were analyzed using Linguistic Inquiry and Word Count (LIWC), a widely used computerized text analysis program that calculates the degree to which psychologically meaningful word categories are present in each corpora of text (Pennebaker *et al.*, 2015a, b). For each word that appears in a response, LIWC identifies the linguistic dimension(s) it belongs to using an extensive internal dictionary. It then calculates the percentage of words in each response that represents a given linguistic dimension (function words, cognitive processing words, etc.). LIWC provides valid and reliable measures of cognitive processing dimensions (Pennebaker *et al.*, 2015b).

First, each response was submitted through LIWC's cognitive processes analysis to obtain a measure of cognitive load. LIWC's analysis of cognitive processes quantifies the percentage-use of cognitive language within each response such that greater usage of cognitive language reflects greater active cognitive processing of topics and events (Brownlow *et al.*, 2019). This measure can be broken down into six sub-measures of cognitive language that represent engagement in a psychologically meaningful cognitive process. The percentage-use of these sub-measures was calculated for each response to obtain measures of the specific mental operations that underlie cognitive load.

Thinking style can also contribute to heightened cognitive load. Analytic thinking is a cognitively-taxing style of thinking that "prompts more careful information processing, thereby increasing attention on content and evidence" (Swami *et al.*, 2014, p. 574). It involves solving problems by breaking down complex concepts into more manageable parts and connecting them through logic and evidence. In other words, analytical thinking reflects formal and logical thinking rather than more informal, intuitive thinking. Although greater experience and expertise can increase the tendency to engage in intuitive thinking (Peters, 2012), engagement in a particular thinking style also depends on individual tendency and preference for a particular thinking style (Szaszi, 2016). To ensure that differences in cognitive load were not driven by thinking style, the linguistic markers of analytic and intuitive thinking were statistically controlled for in officer responses. Thus, responses from less-experienced officers should contain greater usage of cognitive language than responses from experienced officers—even after controlling for thinking style. LIWC's analytic measure was used to assess the percentage-use of analytical (versus intuitive) language in each response. Greater analytical language usage reflects a greater degree of analytical, logical and complex thinking whereas lower usage reflects more intuitive, narrative thinking.

In all, each of the 45 responses that officers provided (5 scenarios × 3 segments × 3 question types) were scored on (a) overall cognitive processes; (b) the six sub-measures; and (c) analytical language.

Results

Because officers provided responses to questions that were nested within segments and scenarios, this violated the assumption of independent observations and required the use of linear mixed-effects models. Seven separate linear mixed-effects models were conducted, each predicting one of the LIWC scores. Officer skill level (less-experienced, experienced) was entered as a fixed effect because classification as an experienced or less-experienced officer was based on a fixed set of criteria, and we sought to make inferences only between those skill levels. The following variables were entered as covariates: the specific question that each response corresponded to (rapid decision, critical cues and assessment), officer gender, education level, years of police service, military service history, the percentage-use of analytical language and the word count of each response to ensure that any linguistic differences were not driven by description length. Scenarios nested within officers were included as random intercepts [2].

Cognitive language. There was a main effect of skill level such that responses from less-experienced officers contained a more cognitive language (M = 14.25, SE = 0.96) than responses from experienced officers (M = 9.91, SE = 0.99), β = 4.34, SE = 1.56, p = 0.01, indicating that less-experienced officers were under greater cognitive load than experienced officers.

Sub-measures of cognitive language. There was a main effect of skill level such that responses from less-experienced officers (M = 4.58, SE = 0.65) contained more insight language than responses from experienced officers (M = 1.03, SE = 0.66), β = 3.55, SE = 0.93, $p < 0.001$. Additionally, there was a main effect of skill level such that responses from less-experienced officers (M = 2.85, SE = 0.26) contained more discrepancy language than responses from experienced officers (M = 1.77, SE = 0.27), β = 1.08, SE = 0.42, p = 0.01 (Figure 1). Less-experienced and experienced officers did not differ on causation, tentativeness, certainty, or differentiation language.

Study 2

Method

Participants. Participants were recruited from two separate agencies and included 21 less-experienced officers and 20 experienced officers. These agencies were not the same agencies that were reported in Study 1. Less-experienced officers were trainees who recently completed their initial law enforcement training or were in the process of completing it as part of a four-year criminal justice program at a small rural university in the USA. Because less-experienced officers were not yet full police officers, their work experience had been limited to carrying out duties under the supervision of an experienced police officer or observing an experienced police officer on duty. Experienced officers were from a large metropolitan agency in the USA and had at least five years of experience and were serving in an active role such as street patrol, undercover officer, etc.

Procedure of original study. Officers watched nine video scenarios depicting a variety of events that are frequently and infrequently encountered by officers. Six scenarios featured a suspect simulating an attack (i.e. UoF scenarios); the other three scenarios were catch trials in which the suspect calmed down and did not simulate an attack. Officers completed situation-assessment and response-selection exercises after each scenario. After two of the six UoF scenarios, officers completed an additional video-stimulated recall procedure that was audio recorded. Officers only completed the recall procedure for two scenarios due to the time-consuming nature of the stimulated-recall procedure [2]. Officers did not know in advance which scenarios they would be completing the stimulated recalls for.

In the stimulated recall, officers re-watched the scenario and were instructed to say "Stop" whenever they remembered having noticed something and/or a specific action or course of action they thought about pursuing. At these stop points, the experimenter paused the video and verbally asked the officer a series of questions to identify cues that the officer noticed, the inferences they made and the prior knowledge or experience that influenced their thinking [2].

After officers verbally responded to these questions, the experimenter continued playing the video and repeated this process for every stop point until the video ended. The timing and number of stop points varied between officers. Officers also completed additional tasks and questions that were not pertinent to the goals of our current investigation (see Suss, 2013 for full study procedures and materials).

Procedure of current study. Transcripts of each officer's stimulated recall were used to obtain the text responses to the stimulated recall questions. Officer responses were aggregated at each pause point given that a response to a question may provide answers to multiple questions (Tofade *et al.*, 2013).

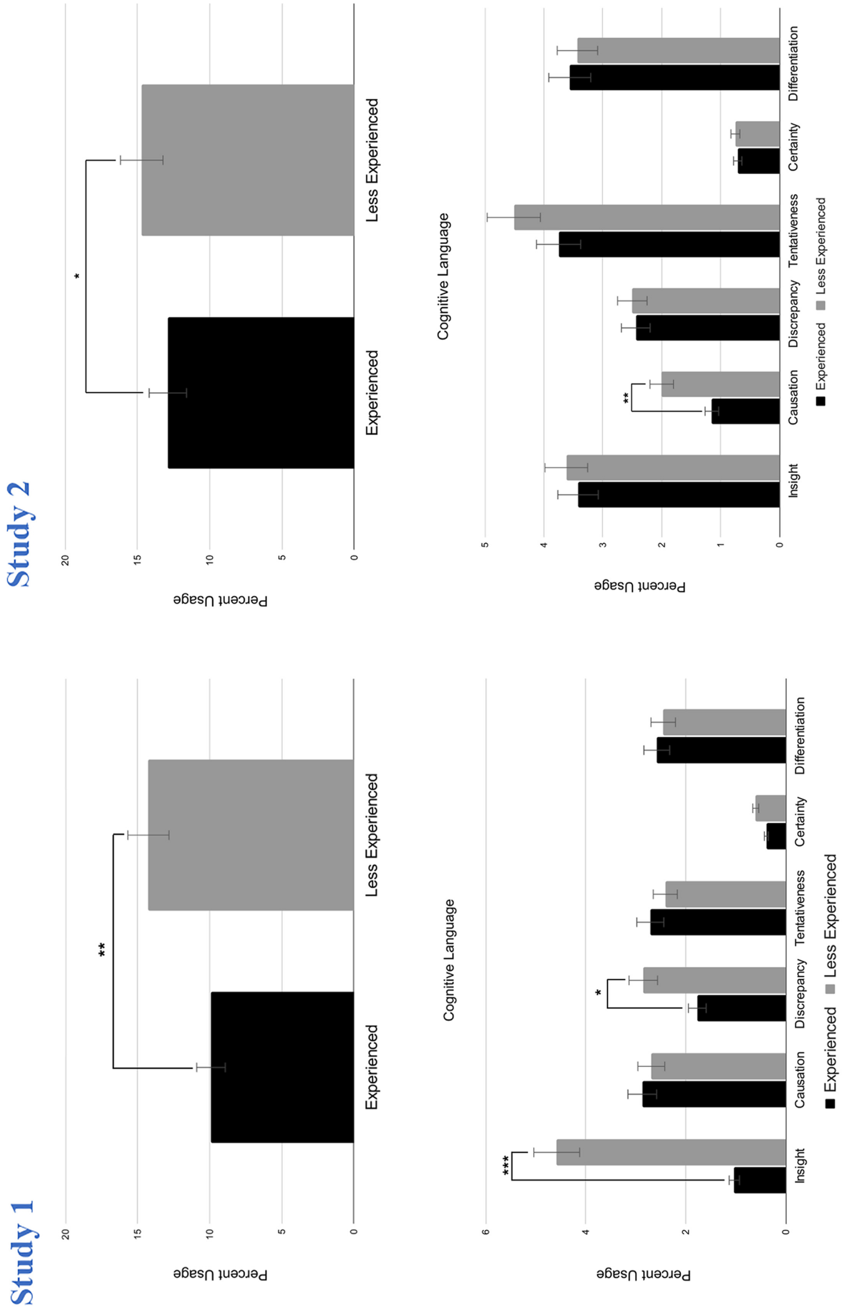

Figure 1.
Comparing usage of cognitive language and sub-measures among experienced and less-experienced officers in Studies 1 and 2

Officers paused around two to six times during each stimulated recall, for a total of 233 pause points across all six scenarios. LIWC was used to score officer responses at each pause point on the following LIWC measures: (a) cognitive processes; (b) the six sub-measures of cognitive processes; and (c) analytical language.

Results

Because officers provided responses to multiple pause points within videos, this violated the assumption of independent observations and required the use of linear mixed-effects models. As in Study 1, seven separate linear mixed-effects models were conducted with each model predicting one of the LIWC scores. Officer skill level (less-experienced, experienced) was entered as a fixed effect for the same reasons stated in Study 1. The percentage-use of analytic language was entered as a covariate. The video, officer and pause point were included as random intercepts [2].

Cognitive language. There was a main effect of skill level such that responses from less-experienced officers contained more cognitive language ($M = 14.70$, $SE = 0.85$) than responses from experienced officers ($M = 12.90$, $SE = 0.85$), $\beta = 1.81$, $SE = 0.86$, $p = 0.04$, indicating that less-experienced officers were under greater cognitive load than experienced officers.

Sub-measures of cognitive language. There was a main effect of skill level such that responses from less-experienced officers ($M = 2.00$, $SE = 0.50$) contained more causation language than responses from experienced officers ($M = 1.15$, $SE = 0.49$), $\beta = 0.85$, $SE = 0.28$, $p = 0.01$. Less-experienced officers and experienced officers did not differ on insight, discrepancy, tentativeness, certainty, or differentiation language (Figure 1).

Discussion

Assessing cognitive load in UoF training is difficult and this is largely due to the methodological issues in existing measures of cognitive load that interfere with instructional procedures and interrupt the training process. Although linguistic measures of cognitive load address these limitations and provide methodological advantages that are not available in other cognitive load measures (Khawaja *et al.*, 2014), it had not yet been applied to police UoF training. To address this shortcoming, officer verbal behavioral data from two separate UoF de-escalation projects were analyzed using a validated natural language processing technique to assess cognitive load between less-experienced and experienced officers. Moreover, the verbal behavioral data were also analyzed to inform the specific thinking patterns that contributed to heightened cognitive load and examine how they differed between less-experienced and experienced officers. This paper demonstrates how a linguistic approach to measuring cognitive load can provide robust, non-intrusive assessments and insight into the cognitive processes that mark expert performance in UoF decision-making tasks. This paper also provides insight into the differences by which less-experienced and experienced officers cognitively process situations involving UoF. These results have practical implications for the process by which police practitioners conduct UoF training and provide many avenues for future research.

Across both studies, responses from less-experienced officers contained greater usage of cognitive language than responses from experienced officers. Consistent with the literature on cognitive load and experience level (Kalyuga *et al.*, 2003), this indicates that less-experienced officers were engaging in more active cognitive processing and were thus experiencing greater cognitive load than experienced officers. Analyzing responses on the six sub-measures of cognitive language offers insight into the specific thinking processes that contributed to heightened cognitive load. In Study 1, responses from less-experienced officers contained greater percentage-use of insight terms and discrepancy terms than responses from experienced officers. This suggested that the heightened cognitive load induced among less-experienced officers during training was particularly characterized by their attempts to evaluate inconsistencies within the scenarios, as well as their attempts to gain a deeper understanding of the incidents in the body-worn camera footage. In Study 2, responses from less-experienced officers contained greater percentage-use of causation terms than responses from experienced officers. This suggests that the heightened cognitive load induced among less-experienced officers was particularly characterized by their attempts to understand what caused the events they observed in the training scenarios.

Although less-experienced officers consistently showed greater cognitive load via greater usage of cognitive language across both studies, the specific thinking processes that differed between less-experienced and experienced officers were not consistent across both studies. This was expected as the specific thinking processes that contribute to heightened cognitive load would depend on the tasks, scenarios and studies at hand. As such, differences in engagement with specific thinking processes between less-experienced and experienced officers should be interpreted within the context of the task, scenario and study at hand. For example, officers in Study 1 were prompted to answer questions about each scenario as they unfolded whereas officers in Study 2 were prompted to answer questions as they re-watched the scenario. Because of this, it is possible that less experienced officers in Study 1 focused more on gaining a better understanding of what was happening in the scenario (insight) and correcting discrepant assessments as the scenario played out (discrepancy) while responses from less experienced officers in Study 2 focused more on explaining why they had engaged in specific thinking processes and connecting it with the events of the scenario (causation). In other words, the specific thinking processes that differed between less-experienced and experienced officers across both studies could have been a function of how officers interacted with the scenarios.

However, the ability to identify the cognitive processes that contribute to cognitive load in specific tasks/situations enables the development of adaptive training interventions. For example, trainers can adjust interventions and target certain mental processes to reduce cognitive load and maximize learning outcomes more efficiently. Put another way, trainers can also use these metrics to adjust a trainee's progression through training. This could be accomplished, for example, by increasing or decreasing scenario or task difficulty or identifying the most appropriate type of question(s) to ask less-experienced officers compared to experienced officers and vice versa to create instructional material that is compatible with—and tailored to—learners' cognitive capacities. This knowledge can also help learners become more aware of their cognitive patterns in specific tasks/situations to regulate their thinking processes more readily and efficiently in real-life UoF situations.

This study has additional practical implications for UoF decision-making training. Linguistic measures of cognitive load are relatively easy to use and do not incur a high labor cost, making it a practical choice for trainers who want to obtain robust, non-intrusive measures of cognitive load without having to invest large sums of money or tackle a steep learning curve. Linguistic measures of cognitive load can be used on existing training programs which maximize training time, does not interfere with established instructional strategies and does not require the restructuring of existing training procedures. The use of linguistic measures of cognitive load can only be used in training programs that feature a component in which officers verbalize (in either written or spoken form) their thinking processes and/or answer questions about the scenarios/tasks at hand. As such, efforts to create new training programs or revise existing ones should consider incorporating a spoken or written verbal component if cognitive load assessment is of interest. Moreover, as natural language processing techniques increasingly become software flexible, linguistic approaches to cognitive load assessment allow for the automatic scoring of trainees on cognitive load in training simulators (as long as trainees provide some form of verbal behavioral data to analyze).

One way in which the measurement of cognitive load could be implemented in UoF training is through computer-based, interactive decision-making training. In this context, measures of cognitive load could be used to implement adaptive training by matching the load of a scenario to the ability of the trainee. In the current studies, we treated scenario as a random effect. However, future studies could use measures of cognitive load as a proxy for determining scenario difficulty. The method described for the current studies could be used to model individual cognitive load over time and in response to properties of scenarios (e.g. type of call, number of suspects). Research on cognitive load during computer-based split-second shoot/do not-shoot decision-making training has highlighted the cognitively demanding nature of such training; increased cognitive load resulted in poorer decision-making performance (Singh *et al.*, 2020). It is likely, then, that excess cognitive load would also have a detrimental effect on performance during other forms of computer-based UoF decision-making training, making it important to match the cognitive load inherent in scenarios to the trainee's current ability.

This paper also provides important contributions to the police training literature. The availability of a robust, non-intrusive and user-friendly method of measuring cognitive load opens the door for more studies to examine the efficacy of CLT-based instruction in UoF training (Mugford *et al.*, 2013). In addition, the specific thinking patterns that contribute to heightened cognitive load are not typically examined in studies assessing cognitive load in UoF decision-making. This information can be used to improve training for less-experienced officers as it provides further insight into the cognitive skills that mark expert UoF decision-making performance. Moreover, the ability to identify the specific thinking processes underlying cognitive load enables police trainers to reinforce training interventions that maximize learning outcomes and facilitate the transfer of skills to the work environment. Given that inappropriate and deadly UoF by police continues to occur despite the substantial resources that have been dedicated to train police officers on making appropriate decisions regarding UoF (Andersen and Gustafsberg, 2016), more research on the improvement of police UoF training is unarguably needed (US Department of Justice, 2020).

More broadly, this paper also contributes to the collective understanding of how less-experienced and experienced officers differ in UoF decision-making. Aside from a relatively small number of studies (Mangels *et al.*, 2020; Ta *et al.*, 2021), few studies have directly compared the UoF decision-making process between less-experienced and experienced officers. This has limited researchers' ability to identify the skills and cognitive processes that encompass expert UoF decision-making. Our results help engender a more complete picture of what comprises expert UoF decision-making. In addition to facilitating training interventions that are consistent with expert UoF performance, this knowledge also provides officers with benchmark expectations for cognitive skill development across their career. In other words, our results help identify the cognitive skills that officers "should" have mastered by the time they reach a particular stage in their career.

This study has several important strengths. LIWC is a tool that provides reliable and well-validated assessments of cognitive load and other psychological states, thus providing robust measurements in the current paper. Because LIWC utilizes a dictionary approach to process natural language, it assesses all responses consistently which provides high concurrent validity (Humphreys and Wang, 2018). By statistically controlling for analytic language in each model, this ensured that differences in cognitive load were not driven by differences in thinking style. The questions officers responded to across both studies are often used to evaluate performance in law enforcement training and are directly pertinent to cognitive load as they ask officers to verbalize their thinking processes as they observed the scenarios.

Greater cognitive load among less-experienced officers was detected across both studies even though (1) the mode by which officer verbal behavioral data were captured differed across studies (written responses in Study 1 versus spoken responses in Study 2); and (2) the data in Study 2 were not collected *in situ* (officers provided their responses after they had already watched a scenario to completion). This also highlights LIWC's validity and flexibility in handling different kinds of verbal behavioral data.

It is possible that differences in cognitive load between less-experienced and experienced officers were driven by departmental factors (such as culture and training) – rather than experience level – given that the samples of less-experienced and experienced officers were drawn from separate agencies in both studies. Although this was mitigated to a degree by the replicated findings across both studies, future research should use representative samples of less-experienced and experienced officers drawn from multiple agencies to control for departmental-level factors and further validate the findings. In fact, we are currently conducting a study that addresses the potential confound between department factors and officer experience. For that study, we are examining whether training interventions can improve adaptive decision making. We recruited sworn officers from three agencies; within each agency, we recruited officers with a wide range of experience. Officers interacted with video-based scenarios in an interactive police judgment-and-decision-making simulator during a pretest and posttest; officers' stimulated recall was elicited following each scenario. After the pretest, officers were randomly assigned to a nine-week online decision-making training program or a no-training control group. Upon completion, this study will yield three sources of verbal/textual data that will be analyzed from a cognitive-load perspective: (1) officers' verbal behavior during the interactive scenarios; (2) officers' textual responses to questions during the online training (i.e. similar to the current Study 1); and (3) officers' verbal responses during the stimulated recall (i.e. similar to the current Study 2).

Additionally, because participants were recruited from US-based agencies, the extent to which our findings generalize to agencies outside the USA is unclear. Country-level variations in police training, culture, organization, funding and community relations may influence cognitive load experienced by less-experienced and experienced officers. Multi-country replication studies should be conducted to assess the external validity of the results. LIWC is not the only natural language processing tool that has been used to obtain valid measures of cognitive load. For example, other studies have used the number of pauses (Müller *et al.*, 2001) and readability (Khawaja *et al.*, 2014) to measure cognitive load. Given that only LIWC was used in the current study, it cannot be determined whether other natural language processing techniques provide a comparable assessment of cognitive load for UoF training. A comparison study should be conducted to identify the natural language processing technique(s) that are most suitable for a given task/scenario in UoF training. Because cognitive load was measured based on officers' responses to similar questions across both studies, future studies should incorporate a wider variety of questions to identify the types of questions that most optimally capture cognitive load.

We did not correlate measures of cognitive load with measures of performance in our study because our focus was on demonstrating the viability of using NLP to identify differences in cognitive load between less-experienced and experienced officers. It is therefore possible that officers with less experience may have implicitly treated the research as a test situation and been more motivated than experienced officers to "get the correct answer." In doing so, they may have expended more cognitive effort when providing their written (Study 1) and verbal (Study 2) responses. Future research should examine the relationship between cognitive load measures and performance to determine whether experienced officers exhibit high levels of performance with less cognitive effort. In doing so, researchers may consider a variety of performance measures. For example, Mangels *et al.* (2020) assessed process performance in terms of officers' use of de-escalatory language. Performance could also be measured in terms of outcome (e.g. correct "shoot/do not-shoot" decisions). In fact, our research team is currently conducting a study in which police officers of varying experience levels interact with simulator scenarios and then reveal their thought processes via video-stimulated recall procedure similar to the one used in Study 2. That study was designed to generate different types of process and outcome measures of performance that could be used to assess the relationship between cognitive load and performance.

In conclusion, this paper provides a promising avenue for the use of officer verbal data to measure and reduce cognitive load in UoF training. The availability of a valid, non-intrusive and practical tool to measure cognitive load allows trainers to actively and efficiently facilitate adaptive UoF training approaches. It also allows researchers to test the efficacy of CLT-based instruction in UoF training more effectively. Moreover, the ability to assess the specific thinking patterns that underlie cognitive load and impair UoF decision-making performance aids the optimization of UoF training interventions, addresses officers' learning needs during training and enhances the transfer of skills to real-life UoF situations. This paper lays the foundation for further investigation into the use of verbal behavioral data to assess officer cognitive load and other important cognitive processes during police training that contribute to adaptive UoF decision-making. Only by constructing a model of expert performance—and contrasting it with models of less-than-expert performance—can the field establish a principled way of developing cognitive aspects of performance in UoF situations. Without such models, instructors are left to rely on their own subjective evaluations of their trainees' performance.

Notes

1. Previous studies using this dataset used the terms *novice and expert* to describe *less-experienced* and *experienced* officers, respectively (Mangels *et al.*, 2020; Ta *et al.*, 2021). Because the data used in Study 2 were derived from a study which referred to officers as *less-experienced* or *experienced,* we also used these terms to refer to *novice and expert* officers, respectively, in Study 1 for uniformity and to avoid confusion.
2. Full models, a breakdown of the number of less-experienced and experienced officers who completed a stimulated recall for each scenario, and the specific questions that officers responded to are reported here: https://osf.io/42p36/?view_only=336125647821474a9eec05ebe048a901

References

Amadieu, F., Van Gog, T., Paas, F., Tricot, A. and Mariné, C. (2009), "Effects of prior knowledge and concept-map structure on disorientation, cognitive load, and learning", *Learning and Instruction*, Vol. 19 No. 5, pp. 376-386.

Andersen, J.P. and Gustafsberg, H. (2016), "A training method to improve police use of force decision making: a randomized controlled trial", *Sage Open*, Vol. 6 No. 2, 2158244016638708.

Arnspiger, B.P. and Bowers, G.A. (1996), "Integrated use-of-force training program", *FBI Law Enforcement Bulletin*, Vol. 65 No. 1.

Ayres, P. and Paas, F. (2009), "Interdisciplinary perspectives inspiring a new generation of cognitive load research", *Educational Psychology Review*, Vol. 21 No. 1, pp. 1-9.

Bennell, C., Jones, N.J. and Corey, S. (2007), "Does use-of-force simulation training in Canadian police agencies incorporate principles of effective training?", *Psychology, Public Policy, and Law*, Vol. 13 No. 1, p. 35.

Beckmann, J.F. (2010), "Taming a beast of burden–On some issues with the conceptualisation and operationalisation of cognitive load", *Learning and Instruction*, Vol. 20 No. 3, pp. 250-264.

Bennell, C., Blaskovits, B., Jenkins, B., Semple, T., Khanizadeh, A.J., Brown, A.S. and Jones, N.J. (2020), "Promising practices for de-escalation and use-of-force training in the police setting: a narrative review", *Policing: An International Journal*, Vol. 44 No. 3, pp. 377-404.

Brownlow, S., Pate, M., Alger, A. and Naturile, N. (2019), “What were they thinking? Analytic and cognitive language in Instagram captions”, *Social Behavior Research and Practice Open Journal*, Vol. 4 No. 1, pp. 21-25.

Dan, A. and Reiner, M. (2017), “Real time EEG based measurements of cognitive load indicates mental states during learning”, *Journal of Educational Data Mining*, Vol. 9 No. 2, pp. 31-44.

De Jong, T. (2010), “Cognitive load theory, educational research, and instructional design: some food for thought”, *Instructional Science*, Vol. 38 No. 2, pp. 105-134.

Humphreys, A. and Wang, R.J.H. (2018), “Automated text analysis for consumer research”, *Journal of Consumer Research*, Vol. 44 No. 6, pp. 1274-1306.

Joseph, S. (2013), *Measuring Cognitive Load: A Comparison of Self-Report and Physiological Methods*, Arizona State University, Tempe, AZ, pp. 1-98.

Kalyuga, S., Ayres, P., Chandler, P. and Sweller, J. (2003), “The expertise reversal effect”, *Educational Psychologist*, Vol. 38, pp. 23-31.

Kennedy, D.O. and Scholey, A.B. (2000), “Glucose administration, heart rate and cognitive performance: effects of increasing mental effort”, *Psychopharmacology*, Vol. 149 No. 1, pp. 63-71.

Khawaja, M.A., Chen, F. and Marcus, N. (2014), “Measuring cognitive load using linguistic features: implications for usability evaluation and adaptive interaction design”, *International Journal of Human-Computer Interaction*, Vol. 30 No. 5, pp. 343-368.

Khawaja, M.A., Chen, F., Owen, C. and Hickey, G. (2009), “Cognitive load measurement from user’s linguistic speech features for adaptive interaction design”, *IFIP Conference on Human-Computer Interaction*, Berlin, Springer, pp. 485-489.

Klein, G., Shneiderman, B., Hoffman, R.R. and Ford, K.M. (2017), “Why expertise matters: a response to the challenges”, *IEEE Intelligent Systems*, Vol. 32 No. 6, pp. 67-73.

Konopasky, A., Durning, S.J., Artino, A.R., Ramani, D. and Battista, A. (2020), “The linguistic effects of context specificity: exploring affect, cognitive processing, and agency in physicians’ think-aloud reflections”, *Diagnosis*, Vol. 7 No. 3, pp. 273-280.

Lemke, J.L. (2012), “Analyzing verbal data: principles, methods, and problems”, *Second International Handbook of Science Education*, Springer, Dordrecht, pp. 1471-1484.

Mangels, L., Suss, J. and Lande, B. (2020), “Police expertise and use of force: using a mixed- methods approach to model expert and novice use-of-force decision-making”, *Journal of Police and Criminal Psychology*, Vol. 35, pp. 294-303.

Moreno, R. (2004), “Decreasing cognitive load for novice students: effects of explanatory versus corrective feedback in discovery-based multimedia”, *Instructional Science*, Vol. 32 Nos 1-2, pp. 99-113.

Mugford, R., Corey, S. and Bennell, C. (2013), “Improving police training from a cognitive load perspective”, *Policing: An International Journal of Police Strategies and Management*, Vol. 36 No. 2, pp. 312-337.

Müller, C., Großmann-Hutter, B., Jameson, A., Rummer, R. and Wittig, F. (2001), “Recognizing time pressure and cognitive load on the basis of speech: An experimental study”, *International Conference on User Modeling*, Springer, Berlin, Heidelberg, pp. 24-33.

Paas, F., Ayres, P. and Pachman, M. (2008), “Assessment of cognitive load in multimedia learning”, *Recent Innovations in Educational Technology that Facilitate Student Learning*, Information Age Publishing Inc., Charlotte, NC.

Paas, F., Tuovinen, J.E., Tabbers, H. and Van Gerven, P.W. (2003), “Cognitive load measurement as a means to advance cognitive load theory”, *Educational Psychologist*, Vol. 38 No. 1, pp. 63-71.

Pennebaker, J.W., Booth, R.J., Boyd, R.L. and Francis, M.E. (2015a), *Linguistic Inquiry and Word Count: LIWC2015 Operator’s Manual*, Pennebaker Conglomerates, Austin, TX, pp. 1-23.

Pennebaker, J.W., Boyd, R.L., Jordan, K. and Blackburn, K. (2015b), *The Development and Psychometric Properties of LIWC2015*, University of Texas at Austin, Austin, TX.

Persky, A.M. and Robinson, J.D. (2017), "Moving from novice to expertise and its implications for instruction", *American Journal of Pharmaceutical Education*, Vol. 81 No. 9, p. 6065.

Peters, E. (2012), "Beyond comprehension: the role of numeracy in judgments and decisions", *Current Directions in Psychological Science*, Vol. 21 No. 1, pp. 31-35.

Ross, D.L. (2000), "Emerging trends in police failure to train liability", *Policing: An International Journal of Police Strategies and Management*, Vol. 23 No. 2, pp. 169-193.

Singh, B., Axt, J., Hudson, S.M., Mellinger, C.L., Wittenbrink, B. and Correll, J. (2020), "When practice fails to reduce racial bias in the decision to shoot: the case of cognitive load", *Social Cognition*, Vol. 38 No. 6, pp. 555-570.

Smith, A.D. and Kelly, A. (2015), "Cognitive processes", in *The Encyclopedia of Adulthood and Aging*, Whitbourne, S.K. (Ed.), doi: 10.1002/9781118521373.wbeaa213.

Suss, J. (2013), "Using a prediction and option generation paradigm to understand decision making", Doctoral dissertation, Michigan Technological University, available at: http://pqdtopen.proquest.com/doc/1413311657.html?FMT=ABS

Suss, J. and Boulton, L. (2019), "Expertise in law enforcement", in Ward, P., Schraagen, J.M., Gore, J. and Roth, E.M. (Eds), *The Oxford Handbook of Expertise*, Oxford University Press, NY, pp. 765-806.

Suss, J. and Ward, P. (2015), in Szalma, J.L. and Scerbo, M.R. (Eds), *Predicting the Future in Perceptual-Motor Domains: Perceptual Anticipation, Option Generation and Expertise*.

Suss, J. and Ward, P. (2018), "Revealing perceptual–cognitive expertise in law enforcement: an iterative approach using verbal-report, temporal-occlusion, and option-generation methods", *Cognition, Technology and Work*, Vol. 20 No. 4, pp. 585-596.

Swami, V., Voracek, M., Stieger, S., Tran, U.S. and Furnham, A. (2014), "Analytic thinking reduces belief in conspiracy theories", *Cognition*, Vol. 133 No. 3, pp. 572-585.

Sweller, J. (1988), "Cognitive load during problem solving: effects on learning", *Cognitive Science*, Vol. 12 No. 2, pp. 257-285.

Sweller, J., Ayres, P. and Kalyuga, S. (2011), *Altering Element Interactivity and Intrinsic Cognitive Load, Cognitive Load Theory*, Springer-Verlag, New York, NY, pp. 203-218.

Szaszi, B. (2016), "The role of expertise and preference behind individuals' tendency to use intuitive decision style", *Journal of Applied Research in Memory and Cognition*, Vol. 5 No. 3, pp. 329-330.

Ta, V.P., Lande, B. and Suss, J. (2021), "Emotional reactivity and police expertise in use-of-force decision-making", *Journal of Police and Criminal Psychology*, Vol. 36 No. 3, pp. 513-522.

Tausczik, Y.R. and Pennebaker, J.W. (2010), "The psychological meaning of words: LIWC and computerized text analysis methods", *Journal of Language and Social Psychology*, Vol. 29 No. 1, pp. 24-54.

Tennyson, R.D. and Volk, A. (2015), "Learning theories and educational paradigms", in Wright, J.D. (Ed.), *International Encyclopedia of the Social and Behavioral Sciences*, 2nd ed., Elsevier, pp. 699-711.

Tofade, T., Elsner, J. and Haines, S.T. (2013), "Best practice strategies for effective use of questions as a teaching tool", *American Journal of Pharmaceutical Education*, Vol. 77 No. 7, p. 155.

United States Department of Justice (2020), "Hearing twelve: policing culture. President's commission on law enforcement and the administration of justice", available at: https://www.justice.gov/ag/page/file/13194 56/download

Van Merrienboer, J.J. and Sweller, J. (2005), "Cognitive load theory and complex learning: recent developments and future directions", *Educational Psychology Review*, Vol. 17 No. 2, pp. 147-177.

Ward, P. and Williams, A.M. (2003), "Perceptual and cognitive skill development in soccer: the multidimensional nature of expert performance", *Journal of Sport and Exercise Psychology*, Vol. 25 No. 1, pp. 93-111.

Westbrook, A. and Braver, T.S. (2015), "Cognitive effort: a neuroeconomic approach", *Cognitive, Affective, and Behavioral Neuroscience*, Vol. 15 No. 2, pp. 395-415.

About the authors

Dr Vivian Ta-Johnson is an Assistant Professor in the Department of Psychology at Lake Forest College. Vivian Ta-Johnson is the corresponding author and can be contacted at: ta@lakeforest.edu

Dr Joel Suss is an Associate Professor in the Department of Psychology at Wichita State University.

Dr Brian Lande is co-founder and Chief Science Officer at Polis Solutions, Inc.