Anderson and Krathwohl's Two-Dimensional Taxonomy Applied to Task Creation and Learning Assessment

Kelsey Urgo, Jaime Arguello, Rob Capra
School of Information and Library Science
University of North Carolina at Chapel Hill
[kurgo,jarguello,rcapra]@email.unc.edu

ABSTRACT

Search tasks play an important role in the study and development of interactive information retrieval (IIR) systems. Prior work has examined how search tasks vary along dimensions such as the task's main activity, end goal, structure, and complexity. Recently, researchers have been exploring task complexity from the perspective of *cognitive* complexity—related to the types (and variety) of mental activities required by the task. Anderson & Krathwohl's two-dimensional taxonomy of learning has been a commonly used framework for investigating tasks from the perspective of cognitive complexity [1]. A&K's 2D taxonomy involves a cognitive process dimension and an orthogonal knowledge dimension. Prior IIR research has successfully leveraged the cognitive process dimension of this 2D taxonomy to develop search tasks and investigate their effects on searchers' needs, perceptions, and behaviors. However, the knowledge dimension of the taxonomy has been largely ignored. In this conceptual paper, we argue that future IIR research should consider both dimensions of A&K's taxonomy. Specifically, we discuss related work, present details on both dimensions of A&K's taxonomy, and explain how to use the taxonomy to develop search tasks and learning assessment materials. Additionally, we discuss how considering both dimensions of A&K's taxonomy has important implications for future IIR research.

1 INTRODUCTION

Search tasks are a central component of interactive information retrieval (IIR). As noted by Toms [19], search tasks play two important roles in IIR research. First, they are used in IIR studies to observe behavior and evaluate systems and tools. Secondly, they can be the main object of study. Research on search tasks helps us understand how task characteristics influence users' needs, behaviors, and pre-/post-task perceptions.

A large body of prior work has focused on understanding how search tasks vary along dimensions related to the task's main activity (e.g., searching vs. browsing), end goal (e.g., well-defined vs. amorphous), and structure (e.g., complexity) [16]. Task complexity is an *objective* task characteristic that has been defined from different perspectives (e.g., [6, 7, 15]). One influential perspective of task complexity is through the lens of *cognitive* complexity. A task's cognitive complexity relates to the types (and variety) of mental activities required to complete the task. The idea behind cognitive complexity originated from educational research. Anderson and Krathwohl [1] proposed a *two-dimensional* (2D) taxonomy for characterizing learning objectives, educational exercises, and learning assessment methods. Anderson and Krathwohl's 2D taxonomy (referred to as A&K's taxonomy) was developed as an extension of Bloom's 1D Taxonomy [4]. The goal of the new taxonomy was to give educators a more complete framework for discussing learning objectives, educational activities, and learning assessment materials. In the context of IIR, Jansen et al. [14] (and later Kelly et al. [15]) built upon the cognitive process dimension of A&K's taxonomy to develop search tasks of varying levels of complexity. Building on these task definitions, studies have found that cognitively complex tasks are perceived to be more difficult and impact search behaviors [2, 8, 13–15, 23]. Specifically, cognitively complex tasks are associated with more search activity [2, 8, 13–15, 23], more trial-and-error (e.g., more abandoned queries) [15], greater use of search assistance tools [8], greater diversity of query-reformulation types [21], and greater *divergence* in search strategies adopted by different users for the same task [15].

As described in Section 3, A&K's taxonomy can be viewed as a table of six columns and four rows. Each column denotes a cognitive process (from simple to complex) and each row denotes a knowledge type (factual, conceptual, procedural, and metacognitive knowledge). In this respect, each of the (6×4) cells in the taxonomy denotes a specific cognitive-process/knowledge-type combination. In this conceptual paper, we revisit A&K's 2D taxonomy and argue that IIR research should consider *both* dimensions of this taxonomy in future research. While prior research has successfully capitalized on the cognitive process dimension (e.g., for designing search tasks and investigating their effects) [2, 8, 13–15, 23], the knowledge dimension has been widely ignored. Specifically, we argue that considering both dimensions of A&K's taxonomy can provide three important benefits (described next).

Avoiding Confounds: First, accounting for both dimensions may help us avoid confounding factors when studying search tasks and their effects on users' needs, behaviors, and perceptions. To illustrate, prior work has found that search tasks involving simple cognitive processes (e.g., memorization) require less effort than

tasks involving more complex cognitive processes (e.g., comparison and evaluation). However, in these studies, knowledge types related to tasks were neither fixed nor manipulated. In fact, as described in Section 7, search tasks designed based on A&K's cognitive process dimension have a noteworthy trend—simple tasks have been designed to involve factual knowledge, moderately complex tasks have been designed to involve conceptual knowledge, and highly complex tasks have been designed to involve procedural knowledge. This trend is somewhat problematic considering that the two dimensions in A&K's taxonomy were designed to be *orthogonal*. It remains to be seen whether the types of cognitive processes required by a search task have the same impact on behaviors and outcomes *regardless of the types of knowledge required by the task*. For example, perhaps cognitive processes have a *reduced* effect when dealing with factual knowledge and a *stronger* effect when dealing with metacognitive knowledge. Prior work has yet to investigate search tasks that vary based on the cognitive process dimension while holding the knowledge type constant.

Studying the Effects of Knowledge Type: Learning-intensive search tasks often involve cognitive processes and specific types of knowledge—knowledge about facts, concepts, procedures and/or metacognitive knowledge (i.e., insights about one's own cognition). IIR research has yet to consider how the knowledge type(s) central to a learning-intensive search task may influence users' needs, behaviors, perceptions, and outcomes. For example, perhaps acquiring factual knowledge requires less effort than acquiring conceptual, procedural, and metacognitive knowledge, respectively. Furthermore, the types of knowledge being sought may influence the types of information (and information sources) most relevant to the task. For example, certain information sources (e.g., Wikipedia) may be more useful for acquiring conceptual knowledge and others (e.g., social Q&A sites) may be more useful for acquiring procedural knowledge. Similarly, certain types of media (e.g., videos) may be particularly useful for acquiring specific types of knowledge (e.g., procedural). If knowledge types impact search behaviors *and* relevance criteria, then this is an opportunity for systems to promote specific types of content based on a user's search behaviors.

Developing Learning Assessment Materials: In recent years, the "search as learning" community has proposed that search can (and should) be viewed as a learning process. In this respect, knowledge gain (or lack thereof) can be viewed as an important metric for evaluating the quality of a user's interaction with a system. Studies in this area have aimed at understanding how search behaviors correlate with knowledge gains [10–12]. In this conceptual paper, we describe how A&K's taxonomy can be used to develop materials to assess learning. Specifically, we describe how the taxonomy can be used to develop post-task exercises and qualitative coding schemes to analyze materials (e.g., essays, summaries) that participants may be asked to produce as part of a study. Furthermore, we argue that knowledge type is likely to play a central role in this process. For example, we hypothesize that acquiring metacognitive knowledge may be a stronger signal of a positive search experience as compared to acquiring merely factual knowledge.

2 RELATED WORK

In this conceptual paper, we build on three areas of prior work. First, we build on prior work that has studied task complexity from different perspectives. Second, we build on prior work on *cognitive* complexity and its effects. Finally, we build on IIR studies that have aimed at measuring knowledge gains during search.

Perspectives on Task Complexity: Prior work has characterized search tasks along many different dimensions. Li and Belkin [16] proposed a unifying framework for characterizing search tasks based on generic facets (e.g., self-motivated vs. assigned) and common attributes—subjective attributes (e.g., perceived difficulty) and objective attributes (e.g., complexity). Task complexity is an important attribute that has been studied from different perspectives. As noted by Wildemuth et al. [20], prior work has treated complex tasks as involving: (1) multiple steps; (2) multiple concepts or concept-types; and/or (3) greater uncertainty. Campbell [7] defined complex tasks as having: (1) multiple outcomes, (2) multiple paths to the outcomes, (3) greater interdependence between the paths, and (4) greater uncertainty about the paths. Prior work has also studied task complexity through the lens of *a priori* determinability—the degree of uncertainty about the task's requirements, form of the solution, and processes involved. Byström and Järvelin [6] used the principle of *a priori* determinability to organize tasks into five complexity levels. On one extreme, *automatic processing tasks* are completely determinable. On the other extreme, *genuine decision tasks* are completely unstructured—neither the task requirements, form of the solution, nor processes involved are known in advance. Bell and Ruthven [3] used the principle of *a priori* determinability to design tasks for a lab study. Complex tasks were designed to have more uncertainty regarding three search-related processes: (1) deciding what information is needed, (2) finding valuable sources, and (3) recognizing relevant information.

Prior work has also defined task characteristics *indirectly* related to complexity. Li and Belkin [16] defined a task's *goal* as either specific or amorphous. In this respect, an amorphous task is less *a priori* determinable (i.e., more complex). Similarly, Li and Belkin [16] defined a task's *product* as either factual (i.e., finding a fact), decision-based (i.e., supporting a decision), or intellectual (i.e., generating new ideas). If mapped to A&K's 2D taxonomy, these correspond to the cognitive processes of remember, evaluate, and create.

Cognitive Complexity: In the context of IIR, *cognitive* complexity relates to the types (and variety) of mental activities required to complete a learning-intensive search task. Jansen et al. [14] were the first to use A&K's 2D taxonomy to develop search tasks of varying levels of cognitive complexity. Jansen et al. developed tasks using all six cognitive process levels: remember, understand, apply, analyze, evaluate, and create. Based on participants' search behaviors, search tasks in the two mid-complexity categories (apply and analyze) required the most effort—had more (and longer) queries, more pages visited, and longer sessions. During highly complex tasks (evaluate and create), participants relied more on their prior knowledge and searched mostly to confirm information.

Wu et al. [23] developed 20 search tasks that varied across 4 topics (e.g., health) and 5 cognitive process levels based on A&K's taxonomy (apply was omitted). In a lab study using 10 of these tasks, highly complex tasks required more search activity (e.g., more queries, pages visited, and longer sessions) and were perceived to be more difficult. Additionally, participants reported greater engagement during mid-complexity tasks, possibly because they were neither too easy nor difficult. In a follow-up study using all 20

search tasks from Wu et al. [23], Kelly et al. [15] found similar trends. Additionally, participants' search strategies were more *divergent* during highly complex tasks. Capra et al. [8] used a subset of tasks from Wu et al. [23] to study the effects of task complexity on participants' use of a search assistance tool that displayed search trails from others who completed the same task. Overall, participants made greater use of the tool during complex tasks. Also, task complexity influenced participants' motivations for using the tool—during simple tasks they used it mostly to confirm information and during complex tasks to discover new search strategies.

Learning Assessment: In recent years, the "search as learning" community has considered two basic questions: (1) What factors contribute to learning during search? and (2) What are behavioral measures indicative of learning? To answer these, IIR studies have employed different methods to measure knowledge gains. Collins-Thompson et al. [10] developed task-specific questions for participants to answer after each task. Each task had six open-ended questions (Q1-Q6), which mapped to the six cognitive processes from A&K's 2D taxonomy. For example, a *remember* question required recalling information, while an *analyze* question required comparing alternatives. Participants' responses were scored based on the presence of conceptual (vs. merely factual) knowledge. Results found several trends. First, participants' perceptions of their own learning were highly correlated with the levels of learning measured in their scored responses. Second, easy tasks had more evidence of "low-level learning" (Q1-Q3) while difficult tasks had more "high-level learning" (Q4-Q6). Third, the time spent reading pages was highly correlated with "high-level learning". Gadiraju et al. [12] measured pre-/post-task knowledge using tests with three answer choices: true, false, and I don't know. Two behavioral measures were positively correlated with knowledge gains: time spent reading pages and the presence of complex query-terms.

Wilson and Wilson [22] developed a qualitative coding scheme to measure learning based on summaries written by participants after searching. Three of their codes were designed based on cognitive processes from A&K's taxonomy. The D-Qual code (*understand*) measured the inclusion of relevant, evidence-based statements. The D-Intrp code (*analyze*) measured the degree to which a participant synthesized information in the summary. Finally, the D-Crit code (*evaluate*) measured the presence of comparisons and conclusions. The D-Qual and D-Intrp codes were more effective at detecting knowledge gains. D-Crit was less effective possibly because participants only had five minutes to write summaries.

3 INTRODUCING A&K'S 2D TAXONOMY
In this section, we provide an overview of A&K's 2D taxonomy [1]. In their book [1], A&K provide detailed explanations and sample vignettes on how to use the taxonomy to define learning objectives and design educational activities and learning assessment materials.

Central to a teacher's work is the notion of an *objective*—an implicit or explicit goal of what the teacher wants a student to learn as a result of the teacher's instruction. In the field of education, frameworks have been developed to better articulate and assess these learning objectives. A&K's 2D taxonomy was developed to enable and encourage educators to more clearly articulate the (sometimes vague) learning objectives students are expected to meet at the end of each school year. The taxonomy provides different categories used to group and more *precisely* define learning objectives. Furthermore, the taxonomy was developed to help educators *align* learning objectives, instructional activities, and assessment materials [1].

From the perspective of A&K, learning objectives consist of a verb and a noun. A&K's 2D taxonomy is constructed by intersecting the cognitive process dimension and the knowledge dimension (Table 1). The cognitive process dimension (the verb) consists of *remember*, *understand*, *apply*, *analyze*, *evaluate*, and *create*.

- "*Remember*–Retrieve relevant knowledge from long-term memory.
- *Understand*–Construct meaning from instructional messages, including oral, written, and graphic communication.
- *Apply*–Carry out or use a procedure in a given situation.
- *Analyze*–Break material into its constituent parts and determine how the parts relate to one another and to an overall structure or purpose.
- *Evaluate*–Make judgments based on criteria and standards.
- *Create*–Put elements together to form a coherent or functional whole; reorganize elements into a new pattern or structure" [1].

The knowledge dimension (the noun of a learning objective) is comprised of *factual*, *conceptual*, *procedural*, and *metacognitive* knowledge.

- "*Factual*–The basic elements students must know to be acquainted with a discipline or solve problems in it.
- *Conceptual*–The interrelationships among the basic elements within a larger structure that enable them to function together.
- *Procedural*–How to do something, methods of inquiry, and criteria for using skills, algorithms, techniques, and methods.
- *Metacognitive*–Knowledge of cognition in general as well as awareness and knowledge of one's own cognition." [1]

Each dimension of the taxonomy lies on a continuum. The cognitive process dimension ranges from low complexity (*remember*) to high complexity (*create*). On the other hand, the knowledge dimension ranges from concrete (*factual*) to abstract (*metacognitive*) knowledge. In this respect, A&K note that *conceptual* and *procedural* overlap "...with some *procedural* knowledge being more concrete than abstract *conceptual* knowledge." [1, p.5].

The nouns and verbs in a given learning objective help to categorize the objective into a particular cell in the 2D taxonomy. For example, consider the objective: "The student will learn to distinguish among confederal, federal, and unitary systems of government" [1, p7]. This learning objective is categorized at the intersection of *analyze* and *conceptual* knowledge. The word "distinguish" points to differentiating/discriminating and the word "systems" indicates generalized structures or frameworks. The placement of a specific objective in the taxonomy informs the instructor about the objective. The above objective targets *analyze*/*conceptual*, informing instructors that the objective is relatively complex (higher than *remember* and *understand*) and involves knowledge that is more abstract than concrete *factual* knowledge.

4 SEARCH LEARNING SCENARIOS
As noted previously, A&K created the taxonomy to help educators define learning objectives, design instructional activities, and develop learning assessment materials. The learning objectives specify the goals an educator has for a student, the instructional activities are designed to foster the learning objectives, and the assessment materials are used to determine if the learning objectives were met.

In the context of IIR, we often design tasks to study how users interact with information retrieval systems and to understand their

Table 1: Search Learning Scenario Prompts in 2Dimensional Taxonomy

The Knowledge Dimension	The Cognitive Process Dimension					
	Remember	Understand	3. Apply	4. Analyze	5. Evaluate	6. Create
A. Factual	Recall the tallest building in the world.	Compare the tallest building in the world to the average depth of the ocean.	Use the height of the world's tallest building to measure the average depth of the ocean in units of the tallest building.	Group some of the world's most famous buildings/structures based on their similar/dissimilar heights.	Determine the best unit of length (e.g., feet, meters, football fields) to describe the height of the world's tallest buildings.	Generate a table of the world's 20 tallest buildings using criteria you think are important and useful to readers.
B. Conceptual	Recite the definition of an identity matrix in linear algebra.	Provide an example of an identity matrix.	Using the definition of an identity matrix, verify whether matrix A is an identity matrix of matrix B.	Determine which of the following matrices are identity matrices and, for each, explain why or why not.	Determine which of the following characteristics of matrices are (or are not) necessary to know if matrix A is an identity matrix of B, and for each, explain why or why not.	Create a question to determine if a student knows what an identity matrix is and if they can use it to solve problems.
C. Procedural	Recall the steps of the mergesort algorithm.	Explain mergesort to someone completely new to sorting algorithms.	Run a mergesort on this set of integers: {9,2,6,3,8,1,4,7}.	Distinguish mergesort from heapsort.	Judge whether mergesort or heapsort is most efficient and explain why.	Propose a novel sorting algorithm that is inspired by (not the same as) mergesort.
D. Metacognitive	Identify strategies for retaining information.	Exemplify how you have demonstrated a particular strength from the Gallup Strengths Finder.	Identify memorization strategies and use those techniques that best help you to memorize the first 15 U.S. presidents in order.	Explain your assumptions of individuals whose strengths are analytical, strategic, and futuristic from the Gallup Strengths Finder.	Determine which memorization strategies work well for you and which do not. Explain why some strategies are more effective for you.	Write a brief essay on your personal learning style incorporating the VARK learning style framework.

behaviors. We design tasks with particular *objectives*, the participants in our studies engage in search *activities* (e.g., querying, assessing relevance, reading documents), and then we sometimes *assess* whether learning occurred while the participant engaged in the task. In IIR, such tasks are often referred to as *search tasks* or, when presented with a motivating context, *simulated work tasks* [5].

In this paper, to distinguish between search tasks with specific learning goals versus other types of search tasks, we define *search learning scenarios* (SLSs). The goal of a SLS is to situate an IIR study participant in a context that motivates them to: (1) search for information, (2) engage with information using specific cognitive processes, and (3) have a specific learning objective defined by a cell in A&K's taxonomy. These latter two aspects distinguish SLSs from simple search tasks—SLSs involve searching, but also engaging with information for the purpose of learning. As we define them, search learning scenarios should contain four main components: (1) a motivating contextual scenario, (2) an information need that will require the user to do some searching, (3) an indication of how the user should cognitively engage with the information, and (4) a specification of how the user will be tested or evaluated. For example, consider the following SLS:

> "Imagine that you and some friends recently visited the new One World Trade Center building in New York City. The building is very tall, but you know there are other tall buildings in the world. You and your friends started discussing what was currently the tallest building in the world. Find out what the tallest building in the world is and be prepared to remember it so that you can tell your friends."

This SLS contains a motivating scenario ("imagine that you and some friends visited..."), an information need ("find out what the tallest building in the world is"), an indication of how the user should cognitively engage ("be prepared to remember it"), and a specification of how the user will be tested or evaluated ("tell it to your friends"). In an IIR study, we can measure users' interactions, perceptions, search behaviors, and learning outcomes from performing the task indicated by the SLS.

5 APPLYING A&K'S TAXONOMY
To design SLSs using A&K's taxonomy, we must consider the cognitive process dimension (verb) and the knowledge dimension (noun). Table 1 displays a set of example *prompts* for each combination of cognitive process and knowledge type. The prompts are designed from the perspective of *learning objectives* (i.e., "The searcher will be able to... <do the thing mentioned in the prompt>"), but are also designed to facilitate the creation of search learning scenarios. To illustrate, the remember/procedural prompt is, "Recall the steps of the mergesort algorithm." A search learning scenario based on this prompt could be: "Imagine that you are taking an introductory computer algorithms course. In class today, the instructor asked you to identify and memorize the steps of the mergesort algorithm. The instructor also mentioned that you should be prepared to remember and write down the steps for a quiz tomorrow." Next, we discuss how to use A&K's cognitive process and knowledge dimensions to design prompts exemplified in Table 1. In Section 6, we describe how to use the prompts to develop learning assessment materials.

5.1 The Cognitive Process Dimension
First, we consider a row of Table 1 to illustrate the differences along the cognitive process dimension. We consider *conceptual* knowledge and move left-to-right across the cognitive process dimension (*remember* to *create*). In A&K's taxonomy, conceptual knowledge is concerned with "knowledge of larger, more organized bodies of knowledge (i.e., concepts, principles, models, theories)" [1, p. 42]. Conceptual knowledge focuses on concepts rather than facts, procedures, or metacognitive insights. Note that for all the Table 1 prompts, we assume participants do not already know the answer.

Remember/Conceptual: This type of task requires the searcher to find and remember conceptual-level knowledge. Our example prompt in Table 1 is, "Recite the definition of an identity matrix in linear algebra." This prompt requires the searcher to find a definition of an identity matrix and be able to remember it when asked at a later time. The task involves conceptual knowledge since it

deals with an organized body of information related to linear algebra, matrix properties, and matrix types (as opposed to factual information such as the height of a building).

Understand/Conceptual: Understand-level tasks in A&K's taxonomy involve "constructing meaning", including "interpreting, exemplifying, classifying, summarizing, inferring, comparing, and explaining" [1, p. 68]. Our example prompt in Table 1 is, "Provide an example of an identity matrix." This task involves the same type of conceptual knowledge of identity matrices as the remember/conceptual task, but extends the cognitive component to require not just remembering a concept's definition, but understanding it well-enough to provide an example. Exemplifying is a process at the understand-level in A&K's taxonomy.

Apply/Conceptual: Apply-level tasks require the learner to "carry out a procedure in a given situation" and include processes such as "executing a procedure to a familiar or unfamiliar task" [1, p. 67]. Our example is, "Using the definition of an identity matrix, verify whether matrix A is an identity matrix of matrix B." This task involves conceptual information, but goes beyond the previous tasks (remember, understand) by asking the participant to verify whether one matrix is an identity matrix of another (e.g., by comparing their properties and/or through matrix multiplication). This involves knowing the definition of an identity matrix (remember), understanding it (understand), and applying it in a given situation. In this way, we see that tasks at higher levels of the cognitive process dimension may also require processes at lower levels.

Analyze/Conceptual: Analyze tasks involve "break[ing] material into its constituent parts and determin[ing] how the parts relate to one another" [1, p. 68]. This can include "differentiating, organizing, and deconstructing" [1, p. 68]. Our example is, "Determine which of the following matrices are identity matrices and, for each, explain why or why not." This task involves differentiating, which is an analyze-level cognitive process in A&K's taxonomy.

Evaluate/Conceptual: Evaluate tasks involve "mak[ing] judgements based on criteria and standards" and may involve processes such as "checking, monitoring, critiquing, and judging" [1, p. 68]. Our example is, "Determine which of the following characteristics of matrices are (or are not) necessary to know if matrix A is an identity matrix of B, and for each, explain why or why not." In A&K's taxonomy, making this type of determination is a type of "judgement based on criteria" [1, p. 68] and is at the evaluate-level of the cognitive processes. Similar to previous tasks, the focus on matrices is at the conceptual level, so it is evaluate/conceptual.

Create/Conceptual: Create tasks are at the highest level of the cognitive process dimension and involve "put[ting] elements together... reorganiz[ing] elements into a new pattern or structure" [1, p. 68]. Create tasks may involve "generating, hypothesizing, planning, designing, producing, and constructing" [1, p. 68]. Our example is, "Create a question to determine if a student knows what an identity matrix is and if they can use it to solve problems". This is still a task at the conceptual level, but extends the cognitive process to involve creating something new.

5.2 The Knowledge Dimension
Next, we consider a column of Table 1. For illustration, we consider the *apply* cognitive process and move down the knowledge dimension from *factual* to *metacognitive*. Apply-level tasks require the learner to "carry out or use a procedure in a given situation" [1,

p. 67]. Differences along the knowledge dimension can be viewed as moving from concrete (factual) to more abstract (metacognitive).

Apply/Factual: Factual knowledge focuses on "the basic elements students must know to be acquainted with a discipline or solve problems in it" [1, p. 46] and includes knowledge of "isolated content, terminology, facts, and specific details" [1, p. 42]. Our example prompt for apply/factual is, "Use the the height of the world's tallest building to measure the average depth of the ocean in units of the tallest building." This task involves using factual knowledge (height of the tallest building; average depth of the ocean) to carry out a procedure (measure one thing in terms of the other).

Apply/Conceptual: Already discussed in Section 5.1.

Apply/Procedural: Procedural knowledge is concerned with "how to do something, methods of inquiry, and criteria for using skills, algorithms, techniques, and methods" [1, p.46]. Our example prompt is, "Run a mergesort on this set of numbers." This requires procedural knowledge about the mergesort algorithm and is an apply-level task because it asks the user to actually perform a mergesort. By its nature, procedural knowledge lends itself to being applied. Thus, the apply/procedural combination is fairly common.

Apply/Metacognitive: Metacognitive knowledge is, "knowledge of cognition in general as well as awareness and knowledge of *one's own* cognition" [1, p. 46]. Our example prompt is, "Identify memorization strategies and use those techniques that best help you to memorize the first 15 U.S. presidents in order." This requires metacognitive knowledge because it requires the participant to reflect on which memorization technique they think would work best for them personally. It involves the cognitive process apply because it requires the user to actually apply a memorization technique.

Combinations of Special Note: Table 1 has prompts for every cognitive process and knowledge type (i.e., each cell in A&K's taxonomy). Some cells may seem confusing (or less common) at first glance. For example, create/factual tasks do not necessarily involve "creating facts". Rather, they can involve using factual knowledge as part of a creative process, such as creating a chart of the world's 20 tallest buildings. This involves "reorganiz[ing] elements into a new pattern or structure" [1, p. 68], where the elements are factual (tallest buildings). Also of note, evaluate/factual could involve verifying whether a fact is true or not. However, care must be taken to ensure that the task requires evaluation, not just recognition (see our Table 1 example). Finally, we note that several combinations along the diagonal of Table 1 are natural pairings that occur frequently (e.g, remember/factual, understand/conceptual, apply/procedural).

6 ASSESSMENT OF LEARNING
In recent years, the "search as learning" community has considered two basic questions: (1) What factors influence learning during the search process? and (2) How do search behaviors correlate with learning (or specific *types* of learning)? Addressing either question requires learning assessment. A&K designed the 2D taxonomy to provide specificity and depth of understanding, not only to learning objectives and activities, but also assessment. In the context of IIR, clearly defined learning assessment materials are necessary to precisely measure what a searcher has and has not learned.

IIR studies have used two methods to assess learning: (1) through targeted questions with pre-assigned answers [12] and (2) through

Table 2: Assessment Criteria in the 2D Taxonomy

The Knowledge	The Cognitive Process Dimension					
Dimension	Remember	Understand	3. Apply	4. Analyze	5. Evaluate	6. Create
A. Factual	Recalled a fact	Inferred information from a set of facts	Used fact 1 in conjunction with fact 2 to demonstrate that fact 3 must be true	Grouped items in list based on similar and dissimilar facts	Explained why a particular fact is more useful or helpful than another	Generated a table using facts that were decided to be useful in discriminating among a set of items
B. Conceptual	Recognized a concept	Summarized a concept in own words	Used the definition of a concept to find a solution	Determined and explained why certain items are or are not grouped within a particular concept	Checked that all necessary information is included in definition of a concept	Created assessment to test if concept is understood
C. Procedural	Recalled a procedure	Explained a procedure in own words	Implemented procedure for an intended purpose	Distinguished one procedure from another	Tested and compared performance across two procedures for a similar task	Proposed a novel procedure that serves the same function
D. Metacognitive	Identified strategies for recalling or recognizing information	Deepened comprehension of a topic by reflecting on personal experience	Used the learning strategies that have been identified to be the most subjectively successful	Explained personal opinions, biases, and/or assumptions given new information	Identified why certain learning strategies are more subjectively successful than others	Wrote a reflective essay on subjective learning style

open-ended questions for which responses are scored using qualitative coding schemes [10, 22]. We refer to the first method as *focused* because of its potential to measure depth of learning with respect to *one* cell in the taxonomy. We refer to the second method as *distributed* because of its potential to measure breadth of learning across *multiple* cells. In this section, we describe how A&K's taxonomy can support both types of learning assessment methods.

Assessment Using Targeted Test Questions: A&K's taxonomy can be used to create targeted questions to assess depth of learning with respect to a specific cognitive process and knowledge type. Most prompts in Table 1 can be trivially modified to serve as targeted questions. For example, to measure remember/factual learning, a searcher could be asked: "What is the tallest building in the world?" To measure evaluate/procedural learning, a searcher could be asked a slight modification of the prompt in Table 1: "Which procedure is most efficient: mergesort or heapsort? Explain and justify your answer." As a final example, to measure apply/metacognitive learning, a searcher could be asked: "Which memorization strategy worked best for you when memorizing the first 15 U.S. presidents?"

Assessment Using Coding Schemes: Open-ended questions provide an opportunity to measure breadth of learning. A&K's taxonomy provides a framework to develop qualitative coding schemes to measure learning across the taxonomy based on an open-ended written response. To illustrate, Table 2 shows assessment criteria that could be used to measure learning with respect to each cell in the taxonomy. Different from the prompts in Table 1, the assessment criteria in Table 2 are framed in more general terms. In this respect, these criteria (or slightly modified ones) could be used in the context of different search learning scenarios, topical domains, and open-ended questions. One could imagine using such prompts to score open-ended responses based on evidence of specific types of learning. Consider the following example:

> "You have been taking a college course on influential authors of the Elizabethan era. As a final assignment your teacher has provided the following prompt: Write an essay on the major works and literary contributions of William Shakespeare. Include in your essay which of these contributions you think has had the most impact on the literary world and which of Shakespeare's major works you personally identify with the most."

This prompt could stimulate a response that provides evidence of knowledge gains related to *different* cells in A&K's taxonomy. Such evidence-bearing statements could be identified using criteria similar to those in Table 2. For example, a response may include specific titles and dates of Shakespeare's works (remember/factual). In describing Shakespeare's literary contributions, the searcher may also reference particular literary constructs credited to Shakespeare:

> "Shakespeare's use of iambic pentameter is seen in *Romeo and Juliet* and *Richard III*. Iambic pentameter is a style of verse consisting of one stressed syllable followed by an unstressed syllable. In contrast, the Witches' speeches in *MacBeth* are written in trochaic tetrameter, which begins with one unstressed syllable followed by a stressed syllable."

In this passage, the searcher has demonstrated an understanding of the concept of iambic pentameter by summarizing it in their own words (understand/conceptual), and has also differentiated this concept from a similar one (i.e., trochaic tetrameter) (analyze/conceptual). Responding to the portion of the prompt that asks about personally identifying with Shakespeare's works, the searcher might offer the following response:

> "I have always disliked *Taming of the Shrew* because of the way it portrays female characters and have, in the past, dismissed its significance when considering Shakespeare's literary contributions. After reading *Power and gender in The Taming of the Shrew* I have re-examined these assumptions and thought more carefully of the nebulous nature of the play, much of which is left up to subjective interpretation."

This passage indicates the searcher's self-reflection of personal biases and considers alternative perspectives, indicative of the analyze/metacognitive cell in Table 2. To be brief, we demonstrated only three possible applications of Table 2 for scoring a response to an open-ended exercise. However, there are many more cells that could be evidenced in a response.

7 DISCUSSION & OPPORTUNITIES

In this conceptual paper, we argue that IIR research should consider both the cognitive and knowledge dimensions of A&K's taxonomy in designing *search learning scenarios* (SLSs) and learning assessment materials. Considering both dimensions of the taxonomy has the following implications.

Avoiding Confounds: As described in Section 2, much IIR research has studied how *cognitive* complexity influences searchers' perceptions and behaviors. Studies have found that cognitively complex tasks are perceived to be more difficult and require more

search activity [2, 8, 14, 15, 23]. All this prior research has used tasks based on the cognitive process dimension from A&K's taxonomy. In other words, the knowledge dimension has neither been controlled nor systematically manipulated.

To illustrate, most of the studies above [2, 8, 15, 21, 23] have used tasks developed by Kelly et al. [15], which vary across four topical domains and five cognitive process levels from A&K's taxonomy.[1] The 20 tasks from Kelly et al. [15] have a notable trend—they tend to fall along the diagonal of A&K's taxonomy. Consider the five health-related tasks. The simplest task (remember) is concerned with factual knowledge: How many people in the U.S. have HIV? The three mid-complexity tasks (understand, analyze, evaluate) are concerned with conceptual knowledge: (1) health risks of football, (2) health risks of artificial tanning, and (3) trade-offs between tattoo removal treatments. Finally, the most complex task (create) is concerned with procedural knowledge: an exercise routine.

This trend is somewhat understandable. It seems sensible for cognitively simple tasks to involve isolated *facts*; moderately complex tasks to involve *concepts* (often defined by their relations to *other* concepts); and highly complex tasks to involve procedures. However, as we have shown, there is more to A&K's taxonomy than cells along the diagonal.

Prior work clearly shows that tasks involving complex cognitive processes require more effort. However, without systematically controlling for the knowledge dimension, open questions remain. For example, is this effect *equally* strong across all knowledge types? Or, does the effect grow *stronger* as we go down the knowledge dimension, from factual to metacognitive?

Studying the Knowledge Dimension: By construction, the knowledge dimension in A&K's 2D taxonomy is *orthogonal* to the cognitive process dimension. During a learning-intensive search task, a searcher may want to acquire a specific type of knowledge. Future IIR research should consider whether the type of knowledge being sought has an impact on searchers' needs and behaviors. Below, we argue that task *difficulty* is likely to increase as a searcher aims to acquire factual, conceptual, procedural, and metacognitive knowledge, respectively. Specifically, we argue that task difficulty may increase both in terms of knowledge acquisition and information search. Generally, as we go down the knowledge dimension, knowledge shifts in three ways: (1) from isolated to interconnected, (2) from objective to subjective, and (3) from concrete to abstract.

Factual knowledge may be the easiest to acquire, for several reasons. First, facts are often isolated bits of information with intrinsic value [1, p. 45]. Facts are "isolated" in the sense that their meaning and value can be independent from other facts. Secondly, facts tend to be objective and verifiable. In other words, acquiring factual knowledge may not require synthesizing opinions.

Gaining conceptual knowledge may be more difficult than factual knowledge. First, concepts are often less "isolated" than facts. In other words, understanding a concept (e.g., authoritarianism) may require understanding related concepts (e.g., other forms of government). Moreover, the distinction between related concepts (e.g., authoritarianism vs. totalitarianism) may be highly nuanced. Secondly, some concepts may be highly subjective (e.g., affirmative

[1]https://ils.unc.edu/searchtasksforiir/kelly_ictir2015_paper.pdf

action), in the sense that their definition may vary by source and evolve over time.

Gaining procedural knowledge may also have its challenges. First, some procedures (or algorithms) may have a succinct label (e.g., mergesort), but others may not (e.g., designing a low-impact exercise routine). This may complicate search-related activities such as constructing effective queries and recognizing relevant information. Secondly, some procedural knowledge may also be highly subjective. A&K explain that while procedures often have a fixed product, the procedure itself may be open-ended in terms of which steps to follow, how to follow them, and in what order [1, p. 53]. Thus, gaining procedural knowledge may require engaging with experiential information, assessing credibility, and synthesizing opinions. The open-endedness of procedural knowledge resonates with Campbell's view of task complexity [7]—acquiring procedural knowledge may require considering different paths to an end product and evaluating their uncertainty.

Metacognitive knowledge seems like the most challenging to acquire. First, by definition, metacognitive knowledge is completely subjective. Metacognitive knowledge looks inward and supports how we (as individuals) assess task difficulty, strategize to solve problems based on our knowledge and abilities, and seek to address gaps in knowledge. In this respect, acquiring metacognitive knowledge may require deeply engaging with information (e.g., to determine what we know and do not know).

Finally, as discussed earlier in the paper, the knowledge dimension also exists in a continuum from highly concrete (factual) to highly abstract (metacognitive). This trend also suggests that SLSs may become more difficult as we descend the knowledge dimension. Purely from a learning perspective (ignoring search), Piaget's work in developmental psychology suggests that learning *begins* with concrete knowledge and transitions to abstract knowledge [18]. Along similar lines, research shows that people have more difficulty remembering attributes of abstract versus concrete concepts [17]. Similarly, from a search perspective, prior work suggests that searchers have more difficulty searching for abstract versus concrete concepts. In a study by Capra et al. [9], participants perceived tasks to be more difficult when they were asked to compare alternatives (e.g., products) along an abstract dimension (e.g., ease of use) versus any dimension(s) of their choice. One possible explanation is that abstract dimensions have more varied vocabulary, making it more difficult to construct queries and recognize relevant content. Indeed, Capra et al. [9] performed a qualitative analysis of participants' queries and found that participants often excluded abstract dimensions in their queries.

Understanding How Searchers Traverse A&K's Taxonomy: As a final implication, future IIR research should also consider how searchers traverse the 2D taxonomy based on their prior knowledge and learning objective(s). To illustrate, consider a novice searcher who wants to learn about a specific procedure for the purpose of solving a tangible problem. In this case, the searcher's end goal is to apply procedural knowledge. An important question is: How is this searcher likely to proceed? Are they likely to directly seek information relevant to the "apply/procedural" cell in the taxonomy—an example of someone using the procedure to solve a similar problem? Or are they likely to start elsewhere? And, if so, is this searcher likely to follow a specific path along the taxonomy? For instance,

one could imagine this searcher following a path based on the following sub-goals: (1) understanding factual knowledge about the domain; (2) understanding conceptual knowledge about the domain; (3) understanding the procedure itself; and (4) applying the procedure to solve the problem at hand.

As another example, consider a searcher who wants to compare and evaluate different concepts (e.g., choose the best medication available based on different criteria). Is this searcher likely to directly seek information relevant to the "evaluate/conceptual" cell in the taxonomy (e.g., a table comparing alternative medications)? Or, are they likely to start elsewhere? Again, one could imagine this searcher following a path based on the following sub-goals: (1) understanding concepts related to the domain (e.g., medications available, cost, effectiveness, dietary restrictions, side-effects); (2) understanding factual knowledge (e.g., learning about alternative medications based on different criteria); and (3) evaluating conceptual knowledge (e.g., choosing the best medication).

There are important implications for understanding how searchers traverse A&K's taxonomy based on their goals. For instance, research in this area could inform the design of tools to provide search support or scaffolding based on "typical" paths followed by users. To gain insights about these paths, studies could give participants SLSs with specific learning objectives (i.e., specific to a cell in the taxonomy) and use think-aloud protocols to understand how participants traverse the taxonomy based on their end goals. Ultimately, one could imagine viewing A&K's taxonomy as a *markov chain* with transition probabilities between cells. Going further, if sub-goals influence search behaviors, then one could also imagine using *hidden markov models* to predict a user's end goal (or current sub-goal) based on their behaviors during the search session.

8 CONCLUSION

In this paper, we have argued that A&K's 2D taxonomy is a valuable framework for studying *search learning scenarios* (SLSs)—search tasks with specific learning objectives. A&K's taxonomy characterizes learning objectives based on two orthogonal dimensions: (1) a cognitive process dimension, which describes the types of mental activities related to the learning objective and (2) the knowledge dimension, which describes the type(s) of knowledge involved. Prior IIR research has capitalized on the cognitive process dimension by designing search tasks associated with different cognitive processes and evaluating their effects. However, the knowledge dimension has been largely ignored. In this paper, we have shown how A&K's taxonomy can be used to: (1) design SLSs to use in IIR studies and (2) develop assessment materials that precisely measure the types of learning that occurred during an information-seeking process.

Furthermore, we have argued that considering both dimensions of A&K's taxonomy can provide several important benefits. First, it can help us avoid confounds in IIR studies that examine the impacts of task characteristics on searchers. Future research should consider SLSs designed based on *both* dimensions (vs. only the cognitive process dimension). Secondly, future research should consider the impact of the knowledge dimension on searchers' needs, perceptions, and outcomes (i.e., learning outcomes and others). In other words, genuine SLSs (from real users) are likely to involve both specific cognitive processes and knowledge types. We have argued that search (for the purpose of learning) is likely to become more

difficult as we descend the knowledge dimension, from (isolated, concrete and objective) factual knowledge to (interconnected, abstract, and subjective) meta-cognitive knowledge. Finally, we have argued that future research should study how searchers traverse the taxonomy to address a specific learning objective (defined by a cell in A&K's taxonomy). In other words, how do searchers *sub-divide* a learning objective into smaller sub-goals that can also be defined by cells in the taxonomy? Research in this direction could inform the design of search tools to support and encourage learning.

In closing, A&K's taxonomy was developed as a framework for educators to more *precisely* define learning objectives, activities, and assessments. Here, we argue that it can serve a similar purpose in the context of IIR and "search as learning".

REFERENCES

[1] Lorin W Anderson, David R Krathwohl, Peter W Airasian, Kathleen A Cruikshank, Richard E Mayer, Paul R Pintrich, James Raths, and Merlin C Wittrock. 2001. *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives, complete edition.*

[2] Jaime Arguello. 2014. Predicting search task difficulty. In *ECIR*. Springer, 88–99.

[3] David J. Bell and Ian Ruthven. 2004. Searchers' Assessments of Task Complexity for Web Searching. In *ECIR*. Springer, 57–71.

[4] Benjamin S Bloom et al. 1956. Taxonomy of educational objectives. Vol. 1: Cognitive domain. *New York: McKay* (1956), 20–24.

[5] Pia Borlund. 2016. A study of the use of simulated work task situations in interactive information retrieval evaluations: A meta-evaluation. *Journal of Documentation* 72, 3 (2016), 394–413.

[6] Katriina Byström and Kalervo Järvelin. 1995. Task complexity affects information seeking and use. *IP&M* 31, 2 (1995), 191–213.

[7] Donald J. Campbell. 1988. Task Complexity: A Review and Analysis. *The Academy of Management Review* 13, 1 (1988), 40–52.

[8] Robert Capra, Jaime Arguello, Anita Crescenzi, and Emily Vardell. 2015. Differences in the use of search assistance for tasks of varying complexity. In *SIGIR*. ACM, 23–32.

[9] Robert Capra, Jaime Arguello, and Yinglong Zhang. 2017. The Effects of Search Task Determinability on Search Behavior. In *ECIR*. Springer, 108–121.

[10] Kevyn Collins-Thompson, Soo Young Rieh, Carl C. Haynes, and Rohail Syed. 2016. Assessing Learning Outcomes in Web Search: A Comparison of Tasks and Query Strategies. In *CHIIR*. ACM, 163–172.

[11] Carsten Eickhoff, Jaime Teevan, Ryen White, and Susan Dumais. 2014. Lessons from the Journey: A Query Log Analysis of Within-session Learning. In *WSDM*. ACM, 223–232.

[12] Ujwal Gadiraju, Ran Yu, Stefan Dietze, and P. Holtz. 2018. Analyzing Knowledge Gain of Users in Informational Search Sessions on the Web. In *CHIIR*. ACM, 2–11.

[13] Xiao Hu and Noriko Kando. 2017. Task complexity and difficulty in music information retrieval. *JASIST* 68, 7 (2017), 1711–1723.

[14] Bernard J Jansen, Danielle Booth, and Brian Smith. 2009. Using the taxonomy of cognitive learning to model online searching. *IP&M* 45, 6 (2009), 643–663.

[15] Diane Kelly, Jaime Arguello, Ashlee Edwards, and Wan-Ching Wu. 2015. Development and evaluation of search tasks for IIR experiments using a cognitive complexity framework. In *ICTIR*. ACM, 101–110.

[16] Yuelin Li and Nicholas J. Belkin. 2008. A faceted approach to conceptualizing tasks in information seeking. *IP&M* 44, 6 (2008), 1822 – 1837.

[17] Allan Paivio. 1965. Abstractness, imagery, and meaningfulness in paired-associate learning. *Journal of Verbal Learning and Verbal Behavior* 4, 1 (1965), 32–38.

[18] Jean Piaget and Margaret Cook. 1952. *The origins of intelligence in children*. Vol. 8. International Universities Press New York.

[19] Eliane G. Toms. 2011. Task-based information searching and retrieval. In *Interactive information seeking, behaviour and retrieval*. Chapter 3, 43–59.

[20] Barbara Wildemuth, Luanne Freund, and Elaine G. Toms. 2014. Untangling search task complexity and difficulty in the context of interactive information retrieval studies. *Journal of Documentation* 70, 6 (2014), 1118–1140.

[21] Barbara M. Wildemuth, Diane Kelly, Emma Boettcher, Erin Moore, and Gergana Dimitrova. 2018. Examining the impact of domain and cognitive complexity on query formulation and reformulation. *IP&M* 54, 3 (2018), 433–450.

[22] Mathew J Wilson and Max L Wilson. 2013. A comparison of techniques for measuring sensemaking and learning within participant-generated summaries. *JASIST* 64, 2 (2013), 291–306.

[23] Wan-Ching Wu, Diane Kelly, Ashlee Edwards, and Jaime Arguello. 2012. Grannies, tanning beds, tattoos and NASCAR: Evaluation of search tasks with varying levels of cognitive complexity. In *IIiX*. ACM, 254–257.