University of Alberta

A new tool for measuring individual differences in conceptual structure

by

Emilio Garnet Gagliardi

A thesis submitted to the Faculty of Graduate Studies and Research
in partial fulfillment of the requirements for the degree of

Master of Science

Psychology

Spring 2010
Edmonton, Alberta

Examining Committee

Chris Westbury, Psychology

Christina Gagne, Psychology

Leendert Mos, Psychology & Linguistics

Patricia Boechler, Educational Psychology

This manuscript would not have been possible without the support of my supervisory committee, the Department of Psychology, the students who participated in my experiments, my family and friends, and everyone that has influenced my life. I have only been able to reach as far as I have because of the generous support of so many people – past and present.

Abstract

Implicit concept mapping (iCmap; Aidman & Egan, 1998), measures: (1) the complexity of conceptual activation, and (2) the degree to which integration is internally consistent. These characteristics describe aspects of both Dual Code theory (DCT; Paivio,1986) and of lexical meaning (Johnson-Laird, 1987). Within the DCT literature, two kinds of representations have been proposed, verbal and nonverbal, and in the case of concrete words both kinds of representations will be activated compared to abstract words, which only have a verbal representation. 40 Participants completed Experiment 1, which aimed to assess degree of conceptual change due to learning. The results revealed no change in performance. 120 Participants completed Experiment 2 with a modified task called, progressive concept mapping (proCmap). The results indicated that concrete nouns had greater consistency between trials relative to abstract nouns, whereas abstract nouns had greater complexity. These results provide confirmatory evidence that proCmap is sensitive to information associated with conceptual structure.

Acknowledgements

This manuscript was a long time in the making, maybe too long on some accounts, and it was Chris Westbury's patience and tremendous support that allowed this project to finally come to fruition. A big thank you to Patricia Boechler for agreeing to be on the committee with such little notice, her generosity made this all possible. Christina Gagne always had a smile and open door for me, to throw around ideas about concepts, statistics, and for thinking about these issues in alternative ways. Leo Mos was a personal mentor and always provided his understanding presence over the course of this project. Michael Dawson helped with some of the mathematical analyses that went into this manuscript, his insight brought clarity to these murky details. Dwayne Dickey and Steven Hamblin were also very helpful in providing guidance in understanding some of the mathematical concepts that I needed to develop to see this work through. Brenda Hannon for feedback on implementing the reading comprehension task and analyzing the data. A special thanks to Cyrus Shaoul and Dwayne Dickey for being so patient with me as I learned the complexities and headaches of LaTeX along with their many code snippets and debugging sessions.

Finally, I owe a great deal of gratitude to Lindsay Hoban, Robyn Christenson, and Kristyn Emmerzael who helped run subjects, provide feedback on materials, and directly interacted with the participants to gather responses for my strange experiments and questions. Without all these people, this manuscript would only be a shadow of what it is. My sincere gratitude, respect and thanks to you all.

Acknowledgements

This manuscript was a long time in the making, maybe too long on some accounts, and it was Chris Westbury's patience and tremendous support that allowed this project to finally come to fruition. A big thank you to Patricia Boechler for agreeing to be on the committee with such little notice, her generosity made this all possible. Christina Gagne always had a smile and open door for me, to throw around ideas about concepts, statistics, and for thinking about these issues in alternative ways. Leo Mos was a personal mentor and always provided his understanding presence over the course of this project. Michael Dawson helped with some of the mathematical analyses that went into this manuscript, his insight brought clarity to these murky details. Dwayne Dickey and Steven Hamblin were also very helpful in providing guidance in understanding some of the mathematical concepts that I needed to develop to see this work through. Brenda Hannon for feedback on implementing the reading comprehension task and analyzing the data. A special thanks to Cyrus Shaoul and Dwayne Dickey for being so patient with me as I learned the complexities and headaches of LaTeX along with their many code snippets and debugging sessions.

Finally, I owe a great deal of gratitude to Lindsay Hoban, Robyn Christenson, and Kristyn Emmerzael who helped run subjects, provide feedback on materials, and directly interacted with the participants to gather responses for my strange experiments and questions. Without all these people, this manuscript would only be a shadow of what it is. My sincere gratitude, respect and thanks to you all.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Language is a wonderful tool that people have learned to exploit in order to shape and understand the world around us. With language we build relationships, we establish public spaces, and we set in motion complex goals. The challenge with which we are presented when investigating language is that many different psychological functions, from phoneme detection to lexical access and semantic activation, participate in the production of meaning. As researchers, the practicalities of conducting research on the nature of language requires of us that we focus our investigations on just a few aspects, ignoring the vast array of factors that are nonetheless actively engaged at other levels of the process.

The present manuscript brings together a set of theories and technologies which are intended to provide additional perspective on the mental representation of the meanings of words. The primary focus will be on the nature of concepts, and the functions that concepts serve in the comprehension of meaning. The reason for the emphasis on conceptual structure is to establish a framework for situating a nascent paradigm built to describe and assess conceptual knowledge. In essence, it extends the traditional semantic categorization task by generalizing the number of pairs of words presented simultaneously. This new paradigm brings with it the promise of a new methodology, new visualizations and new ideas about the structure and functioning of the language system – as well as new barriers, and new challenges to overcome.

As such, the experiments carried out within this manuscript are based on an intriguing computer-based concept categorization task called Implicit Concept Mapping (iCmap; Aidman & Egan, 1998; Aidman & Ward, 2002), a task designed to measure two structural characteristics of conceptual structure. The first characteristic describes the complexity or amount of information used to categorize a set of concepts, while the second characteristic describes how consistent the categorizations are from trial to trial, and ultimately block to block.

Aidman and Egan (1998) built the iCmap paradigm to provide instructors with an automated assessment tool capable of describing the quality of the knowledge constructed by students on a given topic, and also to provide each student with a visual aid that affords further discovery learning. The original work and the replication by Aidman and Ward (2002) assessed the validity of the measurements and the usefulness of the iCmap paradigm as a measurement tool of conceptual understanding.

Unfortunately, both investigations failed to demonstrate some of the expected effects despite what appeared to be a genuinely intuitive task. One of the factors that contributed to Aidman and Egan's (1998) results owes itself to the way that they adopted an approach that Komatsu (1992, p.500) had earlier commented on when he said that "psychologists have traditionally equated knowing the meaning of a word with knowing the concept labeled by a word." The consequence was that Aidman and Egan were unable to make predictions or meaningfully interpret their (null) results.

Despite the initial lacklustre results observed in both investigations, there are two reasons that led this author to propose that the iCmap paradigm warrants further attention. The first is that it is very likely that if the paradigm is re-designed to take into account a theory of conceptual structure along with some additional constraints and assumptions about language processing, then the iCmap data will more precisely represent the complexity of the information that underlies conceptual categorization, and the manner in which that information is used.

The second reason that the iCmap paradigm is worth further attention is that it is an attempt to quantify conceptual knowledge as it is experienced. The relations among concepts, and how individuals experience those relations becomes the focus of the investigation, a focus which mirrors the shift by cognitive researchers to view conceptual structure more ecologically (e.g., Gabora, Rosch, & Aerts, 2009).

There is a great opportunity to expand the iCmap paradigm so that it is a useful tool in the hands of psychologists, linguists, and other researchers interested in gaining insight into the underlying nature of how concepts are experienced, and ultimately how the experience of meaning influences behaviour. What is required is a theoretical framework to interpret the measurements created by the iCmap task. Providing a sound theoretical framework to understand the iCmap paradigm is the primary goal of Chapter 1, so that the experiments presented in Chapters 2 and 3 are meaningfully situated. The remainder of this chapter will briefly outline the theoretical basis of conceptual structure adopted within this manuscript by first distinguishing between words and concepts.

The Relation Between Concepts and Words

The theory of conceptual structure presented below makes use of two foundational terms that have crucial differences so it is important to distinguish between them because they are used throughout the remainder of this manuscript. The first term is *concept* and the second is *word*. Although on the surface these may seem to be roughly equivalent terms, they describe different aspects of language. Words have physical (visual or acoustic) representations. Concepts are more difficult to describe because there are no observable characteristics that unequivocally identify them. Concepts are not things, they are a way of encapsulating and talking about our experience of meaning. Consequently, a tremendous amount of research on the nature of concepts has presented topics surrounding the acquisition of concepts (Nguyen, 2007), how concepts are represented psychologically (Hampton & Moss, 2003) and neurologically (Binder et al., 2005), the kinds of concepts (Medin & Lynch, 2000), and more recently on the ecological nature of concepts (Gabora et al., 2009). The result is an amorphous description that is ill-defined and difficult to test. Notwithstanding these complexities, the goal of this manuscript is to develop a new experimental method which attempts to measure the continuous changes in conceptual structure that people experience above and beyond the verbal definitions that underlie the meaning of words.

Part of the challenge in explicating the structure of concepts is that we only have introspective access to their contents and operations, and it is only in language that we are able to express what we experience there. Thus there is a natural tendency to equate a concept with its label (i.e., a word) plus that word's definitional information. What is needed is a psychological theory of lexical meaning that guides our understanding about how words and concepts relate. According to Johnson-Laird (1987), there are several characteristics that can be ascribed to concepts, the first of which is that meaning is inherently difficult to grasp through introspection. Importantly for Johnson-Laird (1987, p.190), "as the processing of speech proceeds from phonology through words to comprehension, it becomes increasingly dependent on inferences based on the social and physical circumstances of the utterance, on knowledge of the situation to which it refers, and on general knowledge." The implication of this characteristic is that some words have empirical definitions while others have conventional definitions, yet these definitions only constitute a portion of the meaning that is activated when using a concept.

Another characteristic ascribed to concepts by Johnson-Laird (1987) concerns the effect that *linguistic context* has on the recognition of spoken and written words.

Context influences the interpretation of words because words are notoriously ambiguous (Johnson-Laird). For example, a single word presented in isolation (e.g., *bank*) will have a slightly different activated meaning compared to when it is presented in a sentence (e.g., *I slept on the bank*) because the context selects constraints that cannot be accessed from the word in isolation. This issue of resolving ambiguity underscores one of the most problematic distinctions of word meaning, that between the *meaning* or *sense* of a word, and the thing in the world that the word refers to, the *referent*. The word *bank* is not an actual bank, nor is it any particular bank. It is rather a frame that allows language users to refer to a particular building or a place.

Johnson-Laird (1987) developed a psychological theory of meaning that situated reference as the means of disambiguating concepts. Inferences about the concepts underlying expressions are based on knowledge of the reference of those expressions. The meanings of words, as far as they are perceptible to language speakers, do not contain information in themselves about how they are able to refer to particular objects and events. It follows that the capacity to refer to things in the world is not a part of the verbal definition of a word, but rather part of the concept that is informed by the verbal definition of that word.

The characteristic that is perhaps most suggestive about the difficulty of equating word meaning with concepts is that people often do not know the complete meaning of the word. For example, we can understand the sentence ‘He has cancer of the pancreas’ despite the fact that our knowledge of cancer and pancreases may be very incomplete. Regardless of how complete their understanding is, people are able to imagine the state of affairs described by the linguistic context, and do not notice the gap in their knowledge unless it is crucial to understanding the context.

The ability to imagine a state of affairs cannot be mediated by purely lexical representations. Interpretations and inferences must be based on a *theory* about what the concept refers to, a theory which according to Johnson-Laird (1987, p.204) selects the "prototypical member of the class that has [some set of] attributes" that have been acquired through learning. This cannot be an image or prototype but is better understood as a mental model.

A mental model, for the present purposes, is defined as a schema with a set of inter-related default values that can be assumed in the absence of information to the contrary. The existence of a schema affords a speaker the ability to make inferences and theories about the current state of affairs. The main point is that when we talk about concepts, we are talking about the way in which the meanings of words are *used*, and there is more going on than simply adding word definitions together. The discussion next turns to two related issues about categories that motivate the experiments to follow.

Categories Are Not Categorical

To begin the discussion on categories, it is important to point out the relationship between *concepts* and *categories*. The most basic description is that a concept specifies the features, functions, and theories around which particular exemplars of that concept cohere, creating a category. A category refers to the set of exemplars that meet the criteria of a concept and against which novel exemplars are compared when deciding whether they are or aren't properly labeled by the same word (Komatsu, 1992). That is to say, concepts are the mental representations of the information used to classify objects and states of affairs into coherent groupings called categories. The basis for deciding whether an exemplar is or isn't a member of a category is the business of a psychological theory of conceptual structure. Broadly speaking, there are two kinds of theories that address the fundamental nature of concepts: similarity- and explanation-based. Similarity-based theories advance the assumption that objects are classified as exemplars of a category because they share attributes with some abstract specification of a category, or with known exemplars of the category.

Similarity-based Theories of Concepts

Within the similarity-based perspective, three main approaches have been advanced in the literature: 1) the classical view, 2) the family resemblance view, and 3) the exemplar view. According to the *classical view*, categories cohere as a result of necessary and collectively sufficient attributes. Although this perspective is at best an idealization, it does appeal to the observation that people have strong intuitions that words have necessary and sufficient conditions, despite the fact that they cannot always articulate those conditions.

Furthermore, within the classical view category membership is regarded as discrete because there are strong constraints on the attributes: either an object is an instance or it isn't. According to this view, the necessary and sufficient attributes, referred to as the *definitional information*, constitute all that is needed to explain how people understand linguistic relations and how they make inferences about and recognize exemplars of a category. Subsequent research has revealed, however, that the boundaries between categories are graded and not distinct, and that people show *typicality effects*, judging some exemplars as being more or less typical of the category than others (Rosch & Mervis, 1975).

According to the *family resemblance view*, categories cohere by virtue of family resemblances among exemplars. Komatsu identified five characteristics associated with this perspective.

The first characteristic is that *typicality* or *prototypicality* is graded. Exemplars with greater family resemblance to a category are judged to be more typical of that category. The second characteristic is that every attribute specified for a concept is shared by more than one exemplar. The implication is that information contained within a concept is an abstraction across exemplars, and that the overlapping networks of shared attributes thus formed hold categories together. The third characteristic is that an exemplar that shares many attributes with other exemplars of a category will have greater family resemblance than an exemplar that shares only a few attributes. The fourth characteristic is the assumption that the attribute weights are independent and combined by adding, resulting in exemplars and non-exemplars of a concept that can be perfectly partitioned by a linear discriminant function.

The fifth characteristic requires elaboration because there are two ways of conceiving of similarity. One kind of family resemblance focuses on the similarity of the exemplars themselves (i.e., the greater in similarity an exemplar is to other exemplars, the greater its family resemblance), and therefore has an extensional emphasis because it makes no assumptions about how the category is represented mentally (i.e., similarity is judged between specific referents in the world that are being compared as similar). The other kind of family resemblance focuses on the similarity between an exemplar and the central tendencies of the category, which places emphasis on the intensional representation. Hence the fifth characteristic of the family resemblance view is that a concept provides a summary of a category in terms of the central tendencies of the exemplars.

There are two challenges to the family resemblance approach of conceptual structure. The first challenge is that with no *a priori* constraint on the nature of similarity shared by the exemplars of a concept, the family resemblance approach has difficulty specifying which similarities count and which do not when it comes to setting the boundaries between categories. One solution is to adopt the assumption that concepts are constrained ecologically, so that certain categories reflect the natural partitionings of objects in the world by our perceptual systems. The second challenge is related to the characteristic of context-dependence. Because the representations described by the family resemblance view are context free, they cannot explain how levels of family resemblance or relevance weights of attributes are affected by context.

Finally, under the *exemplar view*, if one considers that categories cohere around unique exemplars, similarity is judged according to how similar a novel exemplar is to one or more of the exemplars that constitute a category. Thus, categories cohere because their constitutive exemplars are similar to one another in particular ways. The exemplar view can account for contextual effects because the exemplars that are stored in long-term memory depend on context and goals and therefore can account for contextual effects on typicality judgments.

Explanation-based Theories of Concepts

The discussion now turns to two more complex and dynamic approaches to what constitutes category membership, the explanation-based approaches. Explanation-based theories attempt to explain three characteristics of concepts which are difficult to explain by similarity alone. Those characteristics are: 1) the effect of context on resolving ambiguity, 2) the ability to refer to things in the world, and 3) the intensional relations which form the networks of relations within semantic memory.

The first explanation-based approach is the *schema view* (Murphy & Medin, 1985). This view combines the exemplar and family resemblance views into a single structure. A schema is a single structure that captures characteristics of both the family resemblance approach (by storing information that is abstracted across instances) and the exemplar approach (by retaining information about actual exemplars). It provides a uniform method of simultaneously representing information at different levels of abstraction. In addition to functional information, the schema also includes information about the relationships that hold among the attributes of the exemplars of a concept.

The schema view adds three characteristics of conceptual structure to those already discussed. The first is a new operational definition for an attribute, called a *slot*. Every piece of information that is stored about a concept is stored in a slot. Moreover, the schema specifies just which values can and cannot be stored in a slot. Some slots contain attributive information (e.g., *furry*) or probability distributions (e.g., *human males have a range of possible heights*), while others contain references to specific exemplars (e.g., *Sylvester is an exemplar of cat*). The redefinition of an 'attribute' into a 'slot' allows researchers to specify more types of information within the mental representation in order to move beyond simple perceptual propositions.

The second characteristic added by the schema view is that if there is no value associated with a slot, a default value may be inferred. Default values have a low priority relative to slot values, which allows them to be overridden by context, but they can also be context-free or contingent upon the most frequent or average value. The third characteristic added by the schema view, related to the theory of semantic networks, is an explicit description of the relationships among the slots and slot values. By regarding concepts as networks, the schema view draws attention to the fact that schemata actually include information about two kinds of relationships: relations among the slots within a particular concept, and relations between concepts. It is through the relations between concepts that slots and slot values may be inherited from other concepts.

One of the important shifts in attention under the schema view places emphasis on the kinds of relationships that schemata encode, such as functional or causal relationships between concepts. Perhaps more important than shifting focus away from similarity-based perspectives to explanation-based perspectives is the recognition that, concepts are constructed in working memory to participate within a particular context. In the present manuscript that is assumed to be the case. Concepts are not stable representations, but are rather emergent from the application of certain operations, on a base of information from long-term memory.

The second explanation-based approach is the *mental model* proposed by Johnson-Laird (1987). Under this view a schema in long-term memory includes information about how concepts are to be used to construct mental models in working memory, and information about how exemplars of the concept interact with one another and with other objects and forces in the real world. The mental model view highlights the inferential processes inherent in conceptual thinking. Inferences are based on background knowledge and inchoate theories about how the world works, which in turn affect the construction of mental models.

An important distinction that motivates the explication of a mental model is between *ineffable truth conditions* and *verbal truth conditions*. Ineffable truth conditions for Johnson-Laird, Herrmann, and Chaffin (1984) are often called *semantic primitives* in other cognitive theories of meaning. These are the innate and cognitively impenetrable ingredients that are processed to construct meaning. Though unanalysable, ineffable truth conditions play a major role in the construction, modification, and manipulation of mental models. For example, they specify what a default value may be, they determine the constraints on a slot, and they specify the type of relationships slots may enter into. It is through ineffable truth conditions that intentional relations can be evaluated at all. Verbal truth conditions on the other hand relate concepts to objects in the world. Moreover, the verbal definition is what gradually becomes the theory that brings a set of exemplars into coherence.

In similarity-based models exemplars cohere in virtue of possessing similar attributes. In explanation-based models concepts cohere because of the explanation that relates all the schemata, slots, and relationship types together. In addition to how the schemata relate to each other (at the macroscopic level of the concept and the microscopic level of the slot), mental models also contain referential information, such as whether the concept is used referentially to pick out an exemplar, or whether it is used attributively to locate an exemplar to fit a description.

An important argument made by Johnson-Laird (1987) is that the definition of a mental model requires both ineffable truth conditions and verbal truth conditions to be actively and reciprocally engaged. A linguistic utterance has to be translated into an intentional representation - the proposition the utterance expresses - that is by definition cognitively impenetrable. The second level of representation transforms this cognitively impenetrable intentional representation into a model of the state of affairs expressed by the proposition.

Humans can construct models of the world based on information from the perceptual system, from memory, and through imagination in addition to verbal descriptions (Johnson-Laird et al., 1984). Moreover, these models are constantly under revision based on the incoming stream of experiences coming into working memory. The final point that sets mental models apart from the other previously mentioned theories is that mental models have structure that corresponds to the perceived or conceived structure of the state of affairs. To use an example from Johnson-Laird et al., the semantics of the expression “on the right of” specifies the direction to be scanned in order to form a mental model of such assertions as “A is on the right of B”.

In summary, the mental model view of conceptual structure provides a set characteristics and operations that are ascribed to the fundamental nature of concepts such that it is possible to make inferences in the absence of knowledge, to refer to particulars in the world, and account for the way in which context can attune a concept to a particular meaning.

A Sketch of the Internal Structure of Concepts

In the discussion up to this point, attention has been primarily focused on describing those characteristics that a theory of concepts must specify if it is to adequately account for the way in which people use language. Some of the characteristics warrant more elaboration in order to more specifically situate the experiments that follow because the methodology that is developed in the following chapters assesses only a few aspects of conceptual structure, not the entire theoretical specification presented in the theories above.

Within the schema view of concepts, a slot contains the relevant information about some property or relationship. However, a schema also encodes the constraints placed on those slots. Within the mental model view, slots in particular indicate the kinds of relations that a concept may enter into with other concepts. Information represented at this level is referred to as *intensional*, the mental information about the meaning of the word, and how that information is related within a representation and between representations.

According to Johnson-Laird (1987), a complete description of intensional phenomena should explain: the intensional relations between words and expressions, and how those relations yield states of affairs like *synonymy* or *taxonomy*, and the semantic properties of those words that give rise to situations of *ambiguity* (i.e., unresolved constraint selection), *analyticity* (i.e., truth in virtue of meaning), and *self-contradiction* (i.e., falsity in virtue of meaning).

A theory of conceptual structure also needs to account for *extensional relations*, relations between words and the world as human beings experience it. An explication of extensional phenomena is important because meaning is lived through into the world and is used to pick out aspects of the world, not merely to think about the world.

Although a complete explication of these phenomena is beyond the scope of the present manuscript, they are mentioned here as another way of conceiving the difference between words and concepts. Words present a stable and encapsulated way of describing aspects of the world that are associated with particular contexts and behaviors. Concepts, in contrast, are highly dynamic and inter-connected within a rich network of ontologically distinct meanings, such as intensional and extensional information, that provide the mechanisms necessary for language speakers to actively use meanings to make choices, solve problems and engage in imaginative and creative thinking. For example, the meanings of the words *cup* and *cake* do not explicitly specify how to form the compound word *cupcake*. Instead, through the operations and characteristics of concepts, a new concept is formed based on the context.

Having a rough characterization of the underlying structure of conceptual knowledge allows the experiments described in Chapters 2 and 3 to be more thoroughly described and interpreted. The iCmap procedure provides a way of measuring the underlying conceptual structure used to classify concepts along a semantic dimension. The task, which is described in more detail in the next chapter, presents words on a computer screen and asks participants to classify them into groups. The authors of the iCmap paradigm adopted a perspective of conceptual structural that was radically underspecified even compared to the brief introduction presented in the previous section. As such, they were unable to formulate meaningful explanations to account for the results of their experiments. Despite the theoretical and methodological limitations in their work, the iCmap task appears to provide a useful tool for investigating the internal structure of concepts.

Three Goals of the Present Work

The main purpose of the present manuscript is advance the applicability and utility of iCmap. This can be broken down into three goals. The first goal of this manuscript is to apply some of the knowledge from various cognitive and psycholinguistic research to the theoretical specification of the iCmap paradigm. There are many structural variables that contribute to performance during word-based tasks that remained uncontrolled in the original experiments, a state of affairs entirely correctable and addressed here. For instance, Aidman and Egan (1998) selected an arbitrary number of concepts to categorize during a trial (e.g., eight concepts), and they did not consider word length when selecting concepts. Both of these variables have been shown to have significant effects on performance in other language related research, therefore the absence of control on these variables indicates potential sources of error in the design of the initial iCmap procedure.

Another limitation of Aidman and Egan (1998) was that they adopted no theoretical account of concepts in their investigations, but simply placed the structure and operations of concepts in a 'black box'. By implementing a theoretical account of concepts within the iCmap framework, the task can be modified to generate testable hypotheses about conceptual categorization that are capable of shedding light on the organization and operation of the conceptual system used in word–based tasks. If it can be shown that the iCmap paradigm does measure qualities of conceptual structure then a secondary question emerges: How can the task be further refined and validated so that it can generate new experiments, and contribute new perspectives toward our understanding about the structure and functioning of conceptual representations?

The second goal of this manuscript is to build and test the task in an experimental setting to collect evidence about its efficacy. The iCmap task was intended to measure the complexity and the consistency of the conceptual structures activated and stored in working memory when categorizing lists of words presented on a computer screen. The first experiment serves to provide the data necessary to complete the second goal and is discussed in Chapter 2.

The third goal of this manuscript is to modify the procedure developed by Aidman and Egan (1998) to re-calibrate the iCmap methodology with additional characteristics, and to demonstrate that the re-calibrations have significantly improved the precision and quality of the measurements that are collected through the technique. The assumption is that improving the design of the iCmap task and procedure will reveal unique sources of information that can shed light on the nature of conceptual structure, which will be followed by Experiment 1. Whether this assumption holds depends on whether the modifications prove to be sound. The second experiment in Chapter 3 was designed to demonstrate that they do.

Summary

A way of measuring conceptual structure that is sensitive to context and that takes the problems of the dynamic and graded representations of concepts seriously is needed. The iCmap task potentially provides just such a methodology. However, there are concerns of validity that must be addressed in addition to situating the methodology more soundly in the extant research on concepts. In Chapter 2, a brief overview of concept mapping will be presented, and will be followed by Experiment 1. The results from Experiment 1 will be used to motivate a set of refinements that are hoped will suitably address the issue of validity. These refinements are addressed in the second experiment presented in Chapter 3.

Chapter 2

It is necessary to present a theory of concepts in the present manuscript in order to bridge the understanding of conceptual structure from within cognitive psychology with the practical applications of learning theory, which has adopted a concept-centered approach towards explaining the psychology of meaningful learning. Where learning theory researchers are concerned with creating tools and instructional techniques that foster meaningful learning, cognitive psychologists are more concerned with describing the structural and operational characteristics of concepts within a language system, and developing hypotheses about why the language system operates the way that it does. Not surprisingly, the manner in which concepts are implemented by minds lies beyond the scope of the learning theory research program. The consequence of each discipline focusing on different aspects of meaning is that each constructs qualitatively different descriptions of a 'concept'.

By adopting a learning theory perspective, one becomes concerned with the underlying verbal definition of a concept as well as the encyclopedic and background knowledge, so that new knowledge can be meaningfully integrated into the existing networks of concepts. Phrased in this way, learning theory successfully enriches knowledge structures through the operations and characteristics that mental models afford. Therefore, by extension, meaningful learning is achieved through the application of concept mapping, which is intrinsically tied to the mental models whose explanatory inferences bring into coherence the attributes, inter-attribute correlations and exemplars that participate within each concept, and across the domain of knowledge being taught.

The present chapter provides an overview of concept mapping as it was originally conceived by Novak, Gowin, and Johansen (1983), and then presents the work by Trochim (1989) who takes concept mapping from the assessment of *individual* concept maps to *group* concept maps. Within both of these accounts of concept mapping, the scoring procedures that associate particular values to properties of the concept maps are discussed because these tie directly to the experiments below. Some attempt will be made to relate both accounts of concepts together. For even though Novak et al. and Trochim provide definitions for their respective uses of 'concept', the definitions are different from the one advanced in Chapter 1. With that said, the discussion first turns to the traditional concept map scoring procedure, and then a comparison is drawn between that procedure and the procedure adopted by Trochim before finally turning to implicit concept mapping and its scoring procedure.

Concept Mapping

Concept mapping is technique where an individual creates a 2-dimensional diagram outlining his or her understanding of the relationships between and among important concepts within a given domain of knowledge. Novak et al. (1983) demonstrate that concept mapping is a useful technique to facilitate *meaningful learning*, a type of learning inspired by Ausubel's theory of meaningful learning (Ausubel, 1968). Meaningful learning results when an individual consciously and explicitly ties new knowledge to relevant concepts or propositions they already possess. Rote learning, on the other hand, results when new knowledge is arbitrarily incorporated into cognitive structures. What emerged from the work of Novak et al. (1983) was that concept maps serve to promote meaningful learning.

The connection between learning theory and conceptual structure is that, under

Ausubel's theory, a key factor for potential success in meaningful learning is the *framework of relevant concepts and propositions* that an individual possesses. A proposition within a concept mapping perspective is operationalized as *two or more concepts semantically linked* (Novak et al., 1983). Hence the network of mental models activated and integrated within working memory during the learning process directly and indirectly influence the quality of the learning that takes place.

The procedure of concept mapping, generally speaking, begins by supplying individuals with a set of related concepts that they then physically organize by placing the most inclusive, most general concept at the top (of the page or computer display) and successively place less inclusive concepts at lower positions on a hierarchy (Novak et al., 1983). Imposing a hierarchy is the most difficult characteristic in constructing a concept map because the hierarchy depends on the particular 'unit of knowledge' under consideration, where a unit of knowledge is any proposition. As a consequence, the same concept can be located at various different levels in a concept map for different units of knowledge (Novak et al.). It follows from their claim that the propositions linking concepts must also change accordingly, a situation which further highlights the effect of context on the nature of meaning.

Underlying these claims regarding the structure of knowledge is the hypothesis that meaningful learning leads students to organize knowledge hierarchically; however, Freeman and Jessup (2004) argue that networks of meaning are supportive of hierarchical structure, but are not necessarily bound to that type of organization. It is also the case that some topics are not hierarchical in nature, hence the work of Safayeni, Derbentseva, and Cañas (2005) who develop a framework for cyclic concept mapping, a procedure developed to capture functional or dynamic relationships.

Freeman and Jessup (2004) identify many beneficial applications of concept mapping, such as: expressing a conceptualization to others, collaboration in problem-solving, increasing team performance, increasing shared expectations, and shared understanding. Trochim (1989) suggests that concept mapping encourages a group to stay on task, while they express the conceptual framework in their own language. A few of the observed benefits of adopting a technique like concept mapping are improved organizational cohesiveness and morale (Trochim & Linton, 1986).

West, Pomeroy, Park, Gerstenberger, and Sandoval (2000) use concept maps to quantify changes in conceptual framework by comparing the concept maps of resident medical students before and after training in a particular domain. Their results revealed that second- and third-year residents scored significantly higher than first-year residents before instruction, but not after instruction. More importantly they found evidence to suggest that concept mapping assessment reflects expected differences and change in the conceptual framework of resident physicians.

Novak et al. (1983) observed that students (i.e., grade 7 and 8) of any ability level could be successful in concept mapping, and that factors like motivation were more important than academic performance alone. Along with many other related findings, they point out the absence of a correlation between SAT scores and concept map scores, a result which they argue indicates that the two assessment techniques tap different sets of abilities.

Once a concept map is produced however, there still remains the issue of interpreting it and evaluating it. To accomplish that, Novak et al. (1983) have content experts create “ideal maps” and then compare individual maps to the ideal map. One surprising finding from their work was the observation that some of the individual maps were better representations of the underlying conceptual structure for the domain under investigation than the ideal map, a finding which hints at the ecological validity of the technique.

One of the key issues surrounding the successful use of concept mapping indicated by Freeman and Jessup (2004) is the need to elaborate the methods for assessing and measuring concept maps. Several approaches have been developed to measure concept maps and are presented in the next section.

Concept Map Scoring

Individual Maps

Up to this point in the discussion, concept maps have been used to describe an individual's understanding of a topic in relation to an ideal map made by an expert. The task of quantifying concept maps so that different maps can be meaningfully compared requires a scoring procedure which assigns a numeric value to individual concept maps. Novak et al. (1983) adopted a procedure where the ideal map was assigned points for five characteristics:

1. number of relationships
2. hierarchy
3. branching
4. general to specific
5. cross links

The values are then summed together to produce a total score for the ideal map, which can be used to form a ratio of an observed map score to the ideal map score. The ideal map then serves as a template to score individual maps with the possibility, as mentioned above, for individuals to out perform the ideal map. Used in this way, the procedure assigned values to concept maps that later revealed significant differences in performance between mappers of different ability.

The scoring procedure by Novak et al. (1983) remained the predominant method of assigning values to concept maps for many years. A recent scoring procedure by Allen (2006) reduces the number of characteristics to two: the *breadth* and *depth* of a concept map. Allen argues that the traditional scoring procedure leads to a fundamental problem: highly dissimilar maps can end up with similar scores - an issue of no less importance for the iCmap procedure.

Within the framework developed by Allen (2006), individuals possessing greater understanding in a topic will produce maps with a greater number of branches, branches that are also longer on average compared to individuals with rote understanding. Although his scoring procedure reduces the number of characteristics, it still requires an independent scorer to assign the values to the individual maps.

The scoring procedure begins by calculating the average number of branches emanating from non-terminal nodes and then calculates the average branch length. Allen (2006) models the probability of selecting a subset of independent concepts correctly linked to non-terminal nodes within a particular concept map with a truncated Poisson distribution. Importantly, over the course of Allen's research, his data revealed that the variance of branch length was found to be less than the variance for the number of branches, and therefore not well represented by the Poisson distribution, therefore he selected the Conway-Maxwell-Poisson distribution to model the branch length data. By fitting the data to known distributions, Allen places himself in a position to construct maximum likelihood estimates for the population means, which he uses to generate cumulative probability tables for observed pairs of depth and breadth scores.

The well-specified scoring procedure developed by Allen (2006) establishes reliable estimates of the structural properties of a concept map so long as one is willing to adopt his scoring procedure. Before leaving Allen's work, it is worth noting two competing factors that he made explicit because they contribute to the information contained within a concept map score. On the one hand, as the maps become bigger, the sample estimates improve (the Poisson distribution approaches a normal distribution), while on the other hand the task becomes increasingly difficult to administer beyond a certain number of concepts, a situation that places an upper limit on both the number of branches and the average length of the branches that an individual can generate for a particular topic. Allen's observations are insightful because they identified important behaviors of the scores assigned to concept maps, that within an experimental context at least, the values will range from zero to some number much smaller than the possible potential concepts that could be included in the map.

Group Maps

A slightly different approach to the problem of scoring a concept map was developed by Trochim (1989) who set out to describe the information present in a concept map aggregated from a group of individuals. For Trochim, the process of conceptualization was the articulation of ideas, thoughts, or hunches represented in some physical form. More specifically, he adopted *structured conceptualization* (i.e., concept mapping) as a way to describe a sequence of concrete operationally defined steps and which yields a conceptual (pictorial) representation (Trochim & Linton, 1986).

Although Trochim (1989) maintained the process of collecting individual classification scores, he promoted a collaborative approach where the concepts and their linkages were brainstormed and generated collectively. Within this perspective the concept mappers, as opposed to the researchers, were responsible for assigning a rating to each concept along some salient dimension. As a result, concept mapping within Trochim's framework is a highly collaborative process which produces different kinds of observations compared to Novak et al. (1983) and Allen (2006). This difference stems from the purpose of Trochim's procedure that was aimed at a much broader level of understanding because it was intended to be used for evaluation and program planning at an organizational level.

Within the group mapping procedure, information about concept interrelationships is gathered by using an unstructured card sorting procedure adopted from Rosenberg and Kim (1975). There is overlap with the approach developed by Rosenberg and Kim and the Q-technique developed by Stephenson (1953). In the sorting method of Rosenberg and Kim, participants are asked to partition a set of inter-related objects or terms into different groups on the basis of their 'similarity,' 'relatedness,' or 'co-occurrence'. The biggest difference between the two methods pivots on whether the classifications are categorical (i.e., Rosenberg and Kim) or ranked (i.e., Stephenson).

Following Rosenberg and Kim (1975), Trochim prints each concept on separate 3 X 5 index cards and the complete set of concepts is given to a participant. Each participant is instructed to sort the cards into piles that make sense. The results from each participant are put into a matrix with as many rows and columns as there are concepts. The value 'one' is assigned to two concepts placed in the same pile, and the value 'zero' indicates that they were not. There are two important characteristics that emerge from this scoring procedure. The first is a symmetric similarity matrix, and the second is that the diagonal values in that matrix are all set to 'one' indicating that a concept is always placed in a pile with itself. These are important characteristics because they are also incorporated into the iCmap scoring procedure discussed in the next section.

Once the symmetric similarity matrices are computed, they are aggregated across the group to obtain a final, group similarity matrix. In the aggregated similarity matrix, the values indicate how many participants placed a particular pair of concepts in the same pile, and the diagonal values are equal to the number of participants. The final, aggregated matrix is considered the relational structure of the conceptual domain because it provides information about how the participants grouped the statements (Trochim, 1989). A high value indicates that many participants grouped a pair of concepts together which implies that they are similar. A low value indicates that the concept pair was seldom put together into the same pile, which suggests that they are more distinct.

In order to facilitate interpreting the aggregated similarity matrix, Trochim (1989, p.7) adopts multidimensional scaling, a technique that "takes a table of similarities and iteratively places points on a map so that the original table is as fairly represented as possible" and cluster analysis which "represents higher order conceptual groupings of the concepts".

In the first case, the aggregated similarity matrix is submitted to an MDS procedure, that outputs a map of points which represent the set of concepts (selected and sorted by the participants) where distance between points on the map reflects the similarity scores, the greater the frequency between two concept pairs, the shorter distance between them. Importantly for Trochim (1989), the mathematics underlying MDS are more sound than that of cluster analysis, which leads him to adopt the MDS information as a stronger basis for interpreting the inter-relationships among the concepts than the cluster analysis. In regards to cluster analysis, Trochim found that the best clustering algorithm to match the MDS maps was Ward's algorithm.

Before turning to implicit concept maps, Trochim (1989, p.8) has two warnings concerning these multidimensional analytic techniques. The first is that selecting the number of dimensions for the MDS analysis is an important part of the process, "where 2-dimensions typically produce acceptable solutions", however, a researcher could adopted more or fewer dimensions depending on his or her theory or particular data. The second warning concerns the number of clusters to adopt because as he points out, "all hierarchical cluster analysis procedures give as many possible cluster solutions as there are statements" Trochim (p.8). To aid in the decision-making process, Trochim suggests that a researcher err on the side of more rather than fewer clusters.

Implicit Concept Mapping

Implicit concept mapping emerges out of the types of concept mapping research outlined above. Starting from the position that concept maps are useful and effective instructional aids, Aidman and Egan (1998) adopt the direct learner ratings implemented by Trochim (1989) by selecting a computer-based tool that collects direct learner ratings for a set of concepts. The purpose of such a tool is to facilitate teaching the process of traditional concept mapping, whereby the MDS and cluster maps reveal groupings and higher order similarities within an individual's conceptual structure not previously salient to them, and hence the pictorial representations serve as heuristics when first learning the mapping procedure.

For example, by inspecting the 2-dimensional MDS map, an individual might perceive a regularity (e.g., animals grouped tightly around animals, furniture around furniture, and so forth) which triggers an inference to establish an inchoate theory about why the set of concepts behave they way they do. With that inference or theory "in mind" the learner embarks on a traditional concept mapping task for the same set of concepts, selecting nodes and building branches around those initial *implicit* clusters and maps. Aidman and Egan (1998, p.32) propose such a tool extends Trochim's (1989) work by "evaluating individual conceptual representations," and it is that purpose that ties implicit concept mapping to psychological theories of conceptual structure.

Aidman and Egan (1998) selected the computer-based procedure developed by Burmistrov and Shmeliov (1992) to test their hypotheses. Unlike the free-form brainstorming adopted by Trochim, Burmistrov and Shmeliov design each trial to restrict concept categorizations along a single semantic dimension (e.g., similarity), and force participants to perform three choices per trial (i.e., select one concept that is the most similar to, and two concepts that are the most contrastive to a target concept, which is randomly selected from the list). Participants continue to make the three categorizations per trial until each concept has been the target concept once. The procedure essentially "builds-up" a similarity matrix of concept pairings. The concept pairs (target + concept) are determined by classifying the non-target concept into one of three "piles":

1. similar; assigned a score of 1
2. contrastive; assigned a score of -1
3. not rated; assigned a score of 0

At the end of each iCmap procedure, the final data structure represents the concept proximity information for the list of words categorized along a single semantic dimension. Burmistrov and Shmeliov (1992) argue that the data structure has certain structural properties which they label *complexity* and *internal consistency*. Complexity is a description of the unique information contained withing the similarity matrix. Internal consistency describes how symmetrical the pairings are across the entire task. That is to say, when a concept is the target, the participant must classify one of the remaining concepts to be "similar" to it, and when that concept later becomes the target, the symmetrical solution re-assigns the pairing as "similar".

In the first experiment by Aidman and Egan (1998), their primary goal was to assess the utility of the structural properties of the concept proximity information at differentiating students with different levels achievement. Additionally, they were interested in demonstrating whether or not content experts could infer the criteria adopted by a map maker from the MDS and cluster map reconstructions. The results were suggestive. With strong inter-rater reliability (.89) content experts classified student maps into one of four classes: expert, novice, mixed, and other (i.e., no obvious structure).

Finally, an *expert* and a *novice* map (i.e., a map created with the criteria of using the kinds of superficial features a novice might use to classify concepts) were created by an independent content expert, and the student maps was compared to each. Aidman and Egan developed two additional scores to describe the congruency between student maps and the two template maps: *global similarity to expert* (GSE) and *global similarity to novice* (GSN). It must be assumed that these variables were manually scored by the content experts, because no algorithm or formulae are presented that describe the nature of these variables or what data was used in their calculation.

Aidman and Egan (1998) found significant differences when comparing the expertly classified student maps to the novice student maps according to GSN (z=6.08, $p < .01$) and GSE (z=4.50, $p < .01$). There was a greater difference in scores between the expert and novice student maps in relation to the constructed novice map, compared to the constructed expert map. Additionally, students' performance on a formal test showed significant differences whereby students with expert maps scored higher than students with novice or other maps ($M_E = 13.25$; $M_N = 11.41$; $M_O = 11.30$; $p < .01$), followed by students with mixed maps who scored higher than novices and others ($M_M = 12.25$; $p = .04$). Experts did not out perform the mixed group ($p > .05$).

The most important results for the present manuscript were the performances of the complexity and internal consistency scores relative to each other and the formal test. Complexity had a moderate, negative correlation with internal consistency (r = -.57, $p < .01$) which means that as students identified more unique dimensions among the concepts, they were less consistent in their ratings. Internal consistency was weakly correlated with the formal test score (r = .40, $p < .05$), indicating that those students capable of producing a more internally consistent structure also tended to score higher on the formal test. These results also revealed that the number of unique dimensions contained within the proximity data did not provide useful information in differentiating learners of varying skill levels.

Put another way, the complexity score developed by Burmistrov and Shmeliov (1992), argued to measure an aspect of the conceptual structure for the set of concepts, did not correlate with the formal test score, calling its validity into question. Interestingly, Aidman and Egan (1998) offer no comment as to why the variable might be negatively correlated with internal consistency or why it might fail to correlate with the formal test score. This, not surprisingly, lead Aidman to move away from the structural variables of Burmistrov and Schmeliov to begin focusing on the utility of evaluating the MDS and cluster maps.

The second experiment by Aidman and Ward (2002), continued the examination of the iCmap procedure, but shifted the focus from the structural properties of the proximity data to expert-based evaluation of the reconstructed implicit structure revealed by the MDS and cluster plots. In their experiment, Aidman and Ward were again interested in evaluating the efficacy of "[s]tudying individual differences in conceptual structures" (p.36) because it provides "a means for evaluating and assessing knowledge." (p.36). They hypothesized that both the structural and content evaluation measures, derived from the reconstructed maps would vary as a function of student achievement level. They made no mentioned about the failure of the complexity score to correlate with achievement or even about the existing association with internal consistency, presumably because they had no theoretical account from which to discuss those issues.

Aidman and Ward (2002) found no significant differences between the complexity score and the grade groups ($F > 1$), but did find a difference between internal consistency and grade groups ($F(5, 59) = 3.16$, $p < .05$). Post-hoc analysis revealed only that the top students (i.e., 'A' grade students) were different from the remainder of the students. There were additional analyses on the expert ratings of MDS and cluster maps of the students, grouped by grade, which revealed significant differences ($F(5, 59) = 4.55$, $p < .01$ and $F(5, 59) = 5.21$, $p < .01$ respectively).

Aidman and Ward (2002) suggest that the failure to demonstrate differentiation of complexity scores grouped by grade, and only the single difference for internal consistency, was due to the fact that there is little shared variance in learners' knowledge reflected in both concept mapping and formal test scores. The conclusion that they arrived at was that the kind of concept mapping in the iCmap procedure must tap abilities not well measured by formal assessment techniques, namely multiple-choice test scores. Moreover, they reiterate the claim that concept mapping tends to discriminate more effectively between rote and meaningful learning. However, they go no further to illuminate how the algorithms developed to measure complexity and internal consistency might not measure what they were intended to measure.

Interim Summary

We are now in a position to situate the first experiment. In Chapter 1, a psychological theory of meaning was presented to set the stage for thinking about concepts as mental models in working memory. A mental model and its underlying propositional code work in tandem to select meanings that are attuned to the particular linguistic and situational context. The critical claim by Johnson-Laird (1987) was that resolving semantic ambiguity comes from the referential quality inherent in the mental model, where concepts have intentional meaning structures as well as extentional meaning structures, and that when combined with inferences based on background knowledge, a particular meaning is selected that has the capacity to refer to objects in the world.

The concept as mental model perspective adopted within this manuscript is used to situate the work of learning theorists who design and evaluate tasks that empower individuals and groups to express and visually represent their knowledge, and ideally engage in meaningful learning. By making the process concrete through the use of structured conceptualization techniques, learning researchers have been able to assess the quality of the knowledge constructed by participants in a wide range of educational settings.

Concept mapping emerged as the favored tool to promote meaningful learning, and remained a useful technique for many years. The downside of the concept mapping procedure was the need to manually score the concept maps, with the additional cost of having to teach the technique to students in order to use it effectively. Implicit concept mapping was proposed as a means to solve both of those challenges, by automatically scoring conceptual structure, and providing students with visual aids to guide them during their initial concept mapping lessons.

The results of the two experiments assessing the iCmap paradigm revealed that the structural variables do not discriminate between students of all achievement levels. Even more problematic was the proposal by Aidman and Ward (2002) to adopt manual scoring of the reconstructed MDS and cluster plots. By shifting the assessment from the structural properties of the proximity data to the manually scored reconstructed map, Aidman and Ward not only lost a tremendous amount of information normally contained within the traditional concept mapping approach, but implicitly suggested that the utility of the variables developed by Burmistrov and Shmeliov (1992) was limited. With those issues raised, the discussion now turns to the first experiment.

Experiment 1

The motivation for the current experiment was to conduct a pilot study to implement the iCmap paradigm, and collect a set of data to be analyzed and compared to the results of the existing iCmap literature. While that remains the basic purpose, there were many issues with the iCmap research that warrant some attention and that inform several additions to the design of the present experiment.

The first issue was addressed in Chapter 1 by providing a theoretical account of conceptual structure. Within the iCmap procedure, participants are required to make conceptual categorizations, which are operating at a predominantly intensional level of processing because the choices require linguistic meanings to be processed and compared. As a result, a set of mental models are being activated and stored in working memory, which through a series of inferential processes are compared according to the intentional relationship of “similar” or “contrastive”.

The second issue has to do with the task design adopted in both previous experiments. Recall the criticism raised by Allen (2006) that traditional scoring procedures where completely different concept maps can yield similar values. The mathematical descriptions of complexity and internal consistency (discussed below), yield values that behave similarly for a wide range of concept proximity matrices. This is why the complexity score failed to yield significant results. By using the iCmap procedure only once, both Aidman and Egan (1998) and Aidman and Ward (2002) collected a set of arbitrary scores that when compared to each other essentially canceled out any effects of learner achievement, because a random solution could potentially arrive at the same score as an expert solution.

A common technique to increase power is to incorporate a design where participants submit multiple scores on the same variable. The design selected for Experiment 1 is simply to have participants submit a set of scores, once before learning the definitions of the concepts and once after having received instruction. This kind of design, called a pre-post design, has the beneficial outcome of added statistical control imposed by the addition of a second factor, namely, a subject variable, which allows for an additional partitioning of the total variability in both CC and CI scores.

Thus, the following experiment adopts a pre-post design. It is hypothesized that as students learn the definitions of the concepts, the complexity scores will increase due to the increased number of intentional relations being formed. Furthermore, because the mental models gain additional, salient information about the slot relationships and slot values among the concepts, it is more likely that ratings will become more consistent as these values become prototypical. Consequently, it is hypothesized that training will increase internal consistency scores as well.

Finally, given that the present experiment is still assessing the validity and reliability of the structural variables derived from the similarity judgments, it is useful to compare them to other known measures of related cognitive functions. Although there are a tremendous number of psychometric tests available, reading comprehension appeared to be the most likely to share variability in participant performance during each iCmap trial. Where the iCmap measures might not tap the same skills as multiple choice exams, reading comprehension is a skill used by all participants in order successfully complete the iCmap task. Therefore both reading comprehension (measured prior to the experimental task), and scholastic achievement (e.g., the final grade) in a first-year psychology class from which the concepts are selected was also collected for subsequent analyses.

With that, the purpose of Experiment 1 is to first implement the iCmap paradigm developed by Burmistrov and Shmeliov (1992), and second, to extend the experimental procedure to make it more robust with greater control so that it produces more refined information concerning the behavior of the dependent variables. Once the iCmap procedure has been implemented, there are two different analyses that need to be completed. The first is to thoroughly describe the behavior of the dependent variables, complexity, henceforth referred to as cognitive complexity (CC), and internal consistency (IC).

The second analysis is to determine whether the dependent measures are sensitive to learning. If it is possible to demonstrate a significant change in performance on either CC or IC, that would provide evidence that these variables are capable of describing characteristics of the underlying conceptual structures activated and used in service of completing the iCmap task.

The experiment that follows attempts to adhere to the methods and procedures described by Aidman and Egan (1998) and Aidman and Ward (2002). One challenge that emerged in developing and analyzing the software for Experiment 1 was the inadequacy of the information provided in both articles. In both cases, typographic errors and the exclusion of various mathematical and algorithmic details left certain decisions up to guess work and intuition on the part of the author. Consequently, a minor goal of the present experiment is to clarify the experimental design, and data analysis routines so that future research can move forward from a coherent and consistent framework that is more amenable to replication.

Method

Participants

Forty-two University of Alberta undergraduates (20 female and 22 male; M_{Age} = 19.1, SD = 1.89) volunteered to participate in the experiment. All student participants were asked permission for access to their course grade at the completion of the term, and if they did not wish to consent then an alternative activity was provided of equal educational value. Regardless of whether the participant completed the experiment or activity, he or she received partial course credit. Four participants were excluded from the analyses because they did not participate in the post-instruction mapping session.

Materials

Reading comprehension test

Participants were seated in separate rooms at a computer terminal with the Mac OS X 10.4.0+ operating system and 15" LCD computer monitor. Input was collected from a standard keyboard and mouse.

The test of reading comprehension (RC) used in Experiment 1 was taken from

Hannon and Daneman (2001) and was developed with permission as a computer-based application. The software environment used to develop the RC test was Runtime Revolution version 3.1.1. and is available for both OS X and Windows. The RC test is composed of seven blocks, where the first block is used as a practice session leaving six experimental blocks. Each block began by presenting a set of self-paced instructions.

Following the instructions, each block consisted of three sentences that together constituted an experimental paragraph (see Appendix C). The experimental paragraph is designed to contain a set of five target nouns that share particular semantic features. There are three nonsense terms and two real terms. For example:

A NORT resembles a JET but us faster and weighs more.

A BERL resembles a CAR but is slower and weighs more.

A SAMP resembles a BERL but is slower and weighs more.

In this paragraph, NORT, BERL, and SAMP are the nonsense terms, JET and CAR are the real terms, and *speed* and *weight* are the two semantic features. The number of features per paragraph increased every two paragraphs. Thus, the first two experimental paragraphs had two features, the next two had three, and the final two had four features.

After studying the paragraph, participants responded to a set of *true-false* statements about it. Half of the statements were true, and half were false. There were four types of true-false statements. *Text memory* statements tested memory for information explicitly presented in the paragraph; no prior knowledge was required. *Text inferencing* statements tested inferences about information presented explicitly in the paragraph; no prior knowledge was required. *Knowledge access* statements tested access to prior knowledge; no information from the paragraph was required. *Knowledge integration* statements tested integration of prior knowledge with text information. For each true test statement, there was a corresponding false statement that was constructed by reversing the order of the target nouns in the statement. There were two types of knowledge access statements (i.e., low & high) and three types of knowledge integration statements (i.e., low, medium, & high). A total of 276 experimental statements that were presented across the six experimental blocks. See Table 1 for a breakdown of the number of statements for each component subtype.

Table 1
Reading comprehension test statement types and totals

Component	n
Text-based memory	84
Text-based inference	36
Knowledge integration low	24
Knowledge integration medum	36
Knowledge integration high	36
Knowledge access low	36
Knowledge access high	24
Total score	276

Note. Each n is the sum of the true and false statements for that subtype.

Implicit concept mapping

The iCmap task was originally developed as a software application to run on computers running DOS, but was not available for the present experiment. It was therefore necessary to develop a working version of the iCmap task with Runtime Revolution 3.1.1. There were some minor changes made to the iCmap task.

In the original iCmap task design represented in Figure 1, participants were required to drag one of the words from the concept list to the left column to indicate that it was similar in meaning to the header concept. Participants were also required to drag two concepts to the right column to indicate they were contrasting in meaning. Each item from the concept list became the header concept once after which the task was complete.

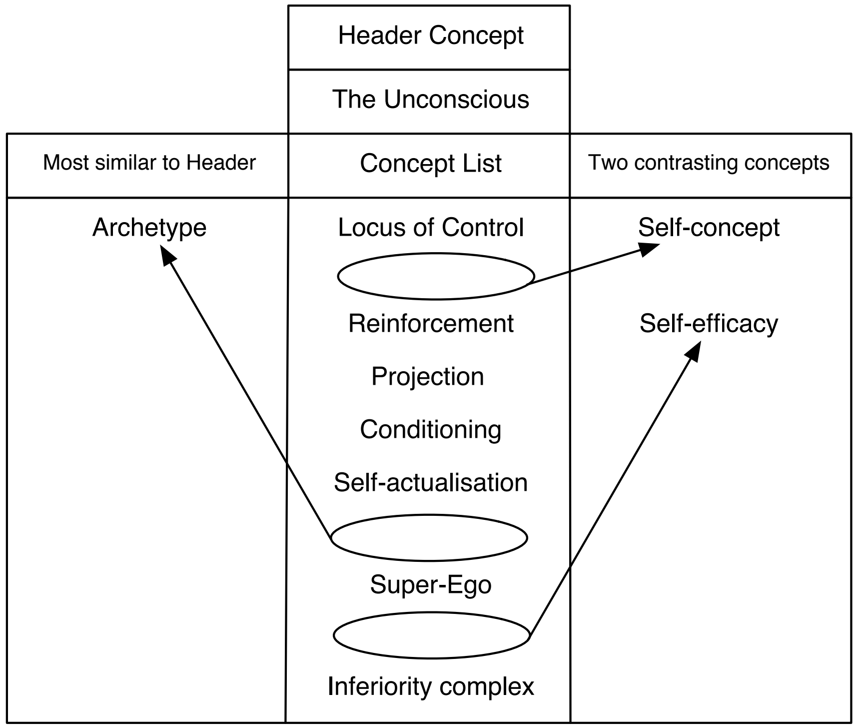

Figure 1
Graphical illustration of the original implicit concept mapping task developed by Aidman and Egan (1998)

The first change eliminated the columnar category assignment mechanism and replaced it with a graphical object called a text field that acted as a bin, which the participant dragged the concepts 'into' (see Figure 2). Furthermore, because the task is a forced choice task, the number of bins was set equal to the number of choices that the participant was required to make.

The second change to the original design was to add a second similar-in-meaning categorization so that participants were required to make two similar categorizations and two contrastive categorizations. There was no justification for an asymmetrical configuration (i.e., one similar vs. two contrastive categorizations), therefore the task was made so that there were equal numbers of judgments for each category.

The final change was to the semantic dimension against which the concepts were classified. The original classification of similar-constrasting was exchanged for related-unrelated. The change was based on the observation that the concept list selected to represent aspects of Operant Conditioning contained many concepts that were operationally related, thus it seemed more sensible to ask participants to judge concepts according to the degree of relatedness rather than which held the most contrasting meaning. Furthermore, because the iCmap task currently contains decisions that require participants to evaluate and compare the intentional information among the mental models, the notion of similarity being so broad might prime some participants to compare the perceptual or functional characteristics of the mental models, which could be superficial, physical features as opposed to more essential intentional information gleaned from the verbal definition.

A side point of interest from the perspective of the author was to build the task to be more intuitive for the participants. Precisely because this is a novel task, any improvements to the usability might also serve to control for variability due to different abilities in learning how to perform it. Hence, a few visual aspects, such as color, font weight, and font family were modified in real-time to emphasize the nature of the task, and how to successfully complete the task. For example, when the participant used the mouse to click on a word, the color of the letters was changed from dark gray to blue to provide explicit feedback that a word had been successfully selected and could be dragged into a bin. Many of the visual design choices were based on the work of Tufte (1991).

The set of concepts that were chosen for Experiment 1 were selected from the two primary textbooks used for introductory psychology courses (Gray, 2002; Carlson, Buskist, Enzle, & Heth, 2005) at the University of Alberta during the semester. Consultation with content experts revealed that the sub-domain of operant conditioning was ideally suited for the present experiment for two reasons: 1) operant conditioning is typically taught half-way through the semester; 2) the key concepts

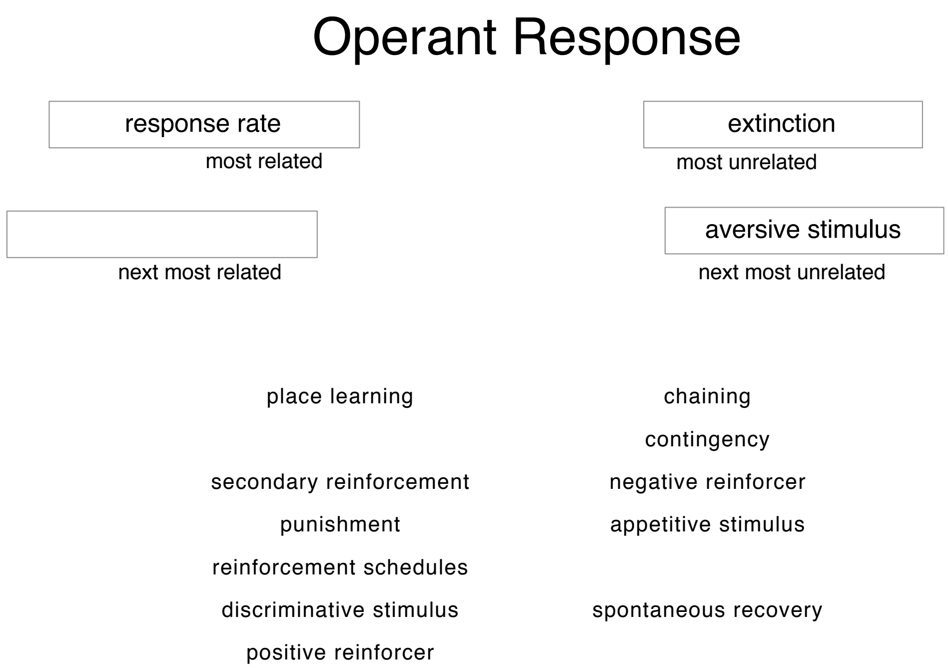

Figure 2
Graphical illustration of the implicit concept mapping task developed for Experiment 1

within operant conditioning are typically well-defined, and, therefore, the kinds of terms found on the multiple choice exams used to assess academic performance.

From an initial list of 30 concepts, sixteen were identified as important by two experienced professors both possessing PhDs in psychology. The final concept list is presented in Appendix A. Prior research by Aidman and Egan (1998) and Aidman and Ward (2002) used eight and eleven concepts respectively.

Procedure

Reading comprehension test

Participants were greeted and given two separate consent forms to read and sign. The first consent form described the nature of the experiment and provided detailed information about their rights while participating in the experiment. The second consent form was required by the Office of the Registrar in order to release the final grades obtained by participants in the introductory psychology course. If the participant did not consent to participate or release their grades they were explicitly assured that they would suffer no academic sanction and were given an alternate activity to complete. The intent of the alternate activity was to provide an educational experience of equal value to that of participating in the experiment. No participant requested the alternative activity.

All the instructions for the RC test were presented on the computer screen, and each participant was also given verbal instructions making explicit reference to the keys on the keyboard for viewing the experimental paragraphs (e.g., a button on the computer screen labeled 'Next Sentence'), and for indicating True and False to the statements (e.g., pressing the 'E' and 'I' keys respectively). Participants were told to read the instructions and proceed at their own pace.

As in Experiments 2 – 4 of Hannon and Daneman (2001), participants were told that they would see a single sentence in the center of the computer screen. Their task was to read the sentence and think about the target nouns and the relations specified between them, and to remember these nouns and relations for the remainder of the block. In particular, they were told that some of the target nouns in the sentence would refer to real-world objects so they should be familiar with them, while the remaining target nouns would represent fictitious objects unfamiliar to them. Each paragraph was composed of three sentences, presented in a randomized order.

The participants both heard and viewed the instructions, which indicated that they should use their real-world knowledge to answer the true-false statements, because some of the statements were considered true based on information contained within the three sentences; other statements should be considered true because the information described in them could be deduced from their existing knowledge about real things in the world.

Participants controlled the length of time that the sentences were displayed, and were told to click on a button labeled 'Next Sentence' to view them. Each sentence was visible only once, and when the third sentence was presented, the 'Next Sentence' button was removed, and the participants were instructed to press the space bar to begin viewing the true-false statements.

After studying each experimental paragraph, participants responded to a series true-false statements presented in random order. While the participants controlled the display time for the experimental sentences, they were told that they only had 12s to respond to each true-false statement once it was presented on the computer screen. The 12s trial duration was chosen by Hannon and Daneman (2001) to provide enough time to respond 'true' or 'false', and also establishing a finite amount of time required to complete the RC test. If a participant did not respond by the end of the trial duration, the statement was scored as an error and the next true-false statement was presented. At the end of each block, participants were presented with a message indicating that they had finished, and the next block commenced.

Participants typically required 25 – 30 min to complete the RC test.

Implicit concept mapping

Participants remained in the same room after completing the RC test. Each participant began the iCmap session with both text-based and verbal instructions indicating that they would be required to read a list of concepts, and that they would have to classify some of them by dragging and dropping them into the bins provided on the computer screen. Each trial of the iCmap task required that participants make two 'related' judgments and two 'unrelated' judgments for each concept that appeared at the top of the screen, called the Header concept.

Specifically, participants were told to read each concept in the list carefully, and select two concepts that were the most related in meaning to the header concept and to drag them into the bins labeled 'related', and to select two concepts that were the most unrelated in meaning to the header concept. When all four bins had words assigned to them, the participant indicated that they were done, and the next trial began. This process repeated until each concept in the list had taken its turn as the Header. The resulting proximity matrix was then analyzed to compute global structural characteristics, cognitive complexity and internal consistency.

Data analysis. The data analysis follows Aidman and Egan (1998) and Aidman and Ward (2002) except for that the DCS-4 program of Burmistrov and Shmeliov (1992) was not used. Instead, the freely available R statistical environment (Forster & Hector, 2002) was used to process all the raw subject proximity data and generate the global structural variables and the associated graphs presented below.

The proximity data were derived for each participant in the form of an $n * n$ non-symmetrical square matrix $A_n = \{a_{ij}\}$, where n = the number of concepts mapped. The cell values were assigned as follows: $a_{ij} = 1$ if the jth concept was judged as related to the ith concept, $a_{ij} = -1$ if the jth concept was judged as unrelated to the

ith concept, and $a_{ij} = 0$ if no judgment was made between the jth and ith concept.

According to Aidman and Egan (1998), the A_n matrix is not a distance matrix because it is not symmetrical and therefore must be transformed in order to compute a factorable metric. Thus, each A_n matrix was transformed into a normalized matrix of all pair-wise scalar products of its rows in a procedure developed and tested by Burmistrov and Shmeliov (1992). The resulting symmetrical matrices S_n were then used in subsequent analyses.

Cognitive Complexity. The first variable computed from each S_n matrix was cognitive complexity (CC), which according to Aidman and Egan (1998) is a variable intended to estimate the number of independent elements, or the degree to which a construct system is broken down. The CC of an S_n matrix represents the number of independent dimensions to which the proximity date can be reduced (Burmistrov & Shmeliov, 1992). It is computed as follows:

$$CC = 1 - \frac{\sum_{i=1}^{n} \sum_{j=1}^{i-1} abs(S_{ij})}{\frac{m \cdot n(n-1)}{2}} \tag{1}$$

Where n = the number of concepts mapped, m = the number of non-zero cells in each row of matrix $\{a_{ij}\}$, and S_{ij} is the scalar product of the i^{th} and j^{th} rows of the matrix:

$$S_{ij} = \sum_{k=1}^{n} a_{ik} \cdot a_{jk} \tag{2}$$

Cognitive complexity is effectively a measure of dimensionality of a proximity matrix, with its values in the range [0,1]. As cited by Aidman and Egan (1998) and Aidman and Ward (2002), CC has been reported to increase as the subject distinguishes more independent properties in the given content domain, with low values indicating a simpler cognitive structure.

Internal Consistency. Internal consistency was estimated by a measure of symmetry of the proximity matrix:

$$IC = 1 - \frac{\sum_{i=1}^{n}\sum_{j=1}^{i-1} SIG_{ij}}{\frac{n(n-1)}{2}} \tag{3}$$

Where $SIG_{ij} = 1$ if sign (P_{ij}) = sign (Q_{ij}), and $SIG_{ij} = 0$ if sign (P_{ij}) $\neq$ sign (Q_{ij}), where P_{ij} is the scalar product of the i^{th} and j^{th} rows and Q_{ij} is the scalar product of the i^{th} and j^{th} columns of $\{A_{ij}\}$. As with cognitive complexity, IC is within the range [0,1], reaching 1 for a symmetrical matrix, which is a perfectly consistent set of relatedness judgments, and diminishing with increased asymmetry of $\{A_{ij}\}$.

Results

Reading comprehension test

The dependent variable for the RC test and its components was accuracy (i.e., number correct). It should be noted that speed of responding (i.e., average reaction time for correct response) was not collected because Hannon and Daneman (2001) obtained results that indicated that speed of responding did not correlate with accuracy and therefore the speed measure likely taps some common factor having to do with speed of reading and responding to a test statement and is not sensitive to the particular component processes that the individual test statements were designed to measure.

The descriptive statistics for the RC test are presented in Table 2. None of the component processes suffered from ceiling or floor effects, and there was an appreciable range of scores for each. The simple correlation matrix of the RC test components is shown in Table 3. The correlations are consistent with the results obtained by Hannon and Daneman (2001). The text-based components, text memory and text inferencing, were highly correlated with one another on both the pre- and post-test (ranging from .83 to .88), while also being very weekly correlated with knowledge access components (ranging from .05 to .22). The knowledge integration components were also moderately correlated with the text-based components (ranging from .36 to .68), each other (ranging from .57 to .71), and the knowledge access components (ranging from .3 to .64).

Table 2
Descriptive statistics for reading comprehension test in Experiment 1

	Pre-test			Post-test		
Component	*M*	SD	Range	*M*	SD	Range
TBM	50.8	8.8	(34−67)	52.1	10.0	(27−67)
TBI	18.3	4.2	(10−25)	19.7	5.1	(6−27)
KIL	16.7	2.6	(9−20)	17.5	2.5	(10−20)
KIM	24.5	3.9	(15−30)	25.7	3.4	(16−30)
KIH	21.7	4.5	(12−30)	23.2	4.9	(12−30)
KAL	26.3	2.0	(20−28)	26.6	1.9	(20−28)
KAH	17.7	2.1	(11−20)	17.7	1.8	(13−20)
Total score	175.9	21.7	(133−216)	182.5	24.5	(117−219)

Note. n = 38. *M* = Mean. SD = Standard deviation.
TBM = Text-based memory; TBI = Text-based inference; KIL = Knowledge integration low; KIM = Knowledge integration medium; KIH = Knowledge integration high; KAL = Knowledge access low; KAH = Knowledge access high.

Due to a programming error, the true-false statements could not be submitted to a reliability analysis because the identifier for each statement was not recorded and the statements were randomized for each participant. However, in light of the congruence between the results obtained here and those obtained in Hannon and Daneman (2001), there appears no reason why the reliability of the RC test (both in terms of the components and the overall test score) would not be similarly reliable.

Implicit concept mapping

There were two dependent measures derived from each proximity matrix: 1) cognitive complexity, and 2) internal consistency. The descriptive statistics for these

Table 3

Simple correlation matrix for reading comprehension test for both pre- and post-test in Experiment 1

Component	2	3	4	5	6	7
1. TBM	.83**(.88)**	.44**(.53)**	.47**(.62)**	.63**(.77)**	.21 (.32)	.22* (.25)
2. TBI		.36* (.48)**	.47**(.49)**	.61**(.73)**	.05 (.37)*	.21 (.33)*
3. KIL			.57**(.73)**	.68**(.60)**	.64**(.54)**	.39* (.60)**
4. KIM				.71**(.59)**	.42**(.56)**	.31 (.54)**
5. KIH					.45**(.38)*	.30 (.43)**
6. KAL						.41* (.52)**
7. KAH						

Note. n = 38. Post-test correlations are in parentheses.
TBM = Text-based memory; TBI = Text-based inference; KIL = Knowledge integration low; KIM = Knowledge integration medium; KIH = Knowledge integration high; KAL = Knowledge access low; KAH = Knowledge access high.
* $p < .05$. ** $p < .01$.

measures are presented in Table 4. CC has the same mean at both sessions as does IC, contrary to the hypotheses that both would increase as a result of learning. This suggests that participants did not add more complexity in their relatedness judgments despite having received instruction covering the definitions for the concepts. Further investigation of the overall behavior of these measures (See Figure 3) indicates three patterns. The first is that IC tends to be greater than CC, which suggests that participants are choosing a response strategy that prefers simple relationships that are easier to recall from trial to trial.

The second pattern that emerged from an inspection of the means indicates that the scores remain unchanged after receiving instruction. The boxplots in Figure 3 demonstrate that neither CC nor IC are sensitive to a change in understanding of the underlying meaning of the concepts used in Experiment 1. The third pattern is that there is very little variability in the scores, which is indicated by the range of CC (.63 to .84) and the range of IC (.78 to .94).

In order to confirm the observation that the means of both measures were un-

Table 4
Descriptive statistics for global structural characteristics for both pre- and post-test in Experiment 1

	Pre-test			Post-test		
Variable	*M*	SD	Range	*M*	SD	Range
Cognitive Complexity	.76	.039	(.67−.82)	.76	.051	(.63−.84)
Internal Consistency	.85	.037	(.78−.94)	.86	.036	(.78−.94)

Note. n = 38. *M* = Mean. SD = Standard deviation.

affected by learning, the scores for both variables were submitted to paired sample t-tests. There was no significant difference for CC (t(37) = -0.35, $p > .05$), or for IC (t(37) = -1.70, $p > .05$). Thus the hypotheses that both CC and IC would increase with learning are not supported.

The correlations between the global structural variables and the component processes of reading indicate that there is no association between an individual's reading ability and the relatedness judgments they generate during an iCmap session (see Table 5). Further, there were no significant correlations between the global structural variables and academic performance or between reading ability and academic performance (see Table 6). There were two significant correlations that emerged from the analyses. First, the RC total scores, pre- and post instruction, were significantly correlated (r = .64) with each other. Second, CC was correlated with itself in the pre- and post-instruction conditions.

Post-hoc analyses. The absence of significant differences between pre- and post-instruction for either CC or IC, and the failure to locate significant correlations between either the RC test, academic performance or the global structural variables, indicates that hierarchical regression of reading ability on either CC or IC is not warranted. Thus the hypothesis that reading ability explains a portion of the variability in CC or IC is not tenable given the present results.

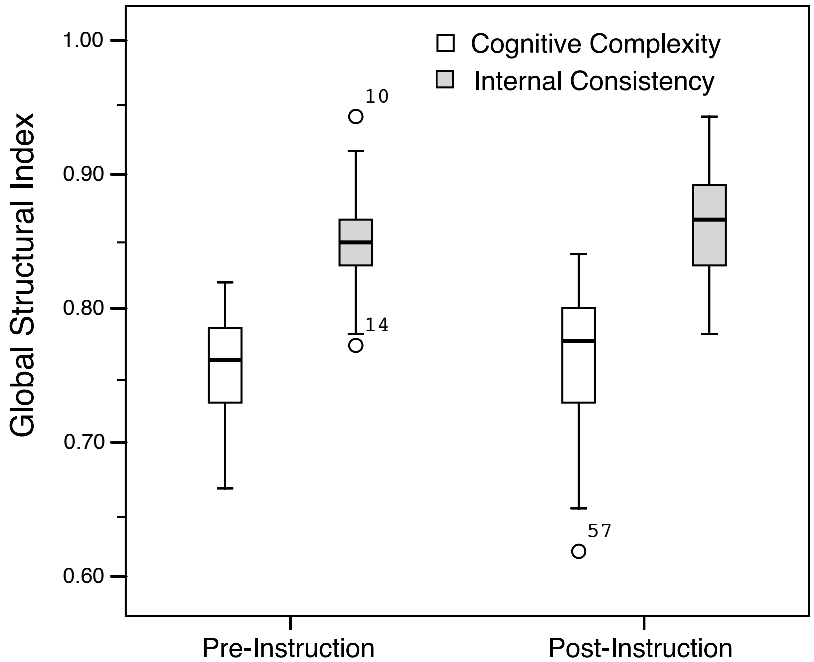

Figure 3
Distribution of scores for both dependent measures for both pre- and post instruction in Experiment 1

Table 5
Simple correlations between reading component processes and global structural characteristics in Experiment 1

	Pre-test		Post-test	
Component	CC	IC	CC	IC
TBM	.08	-.21	.05	-.26
TBI	.14	-.17	.06	-.20
KIL	.00	-.06	.08	-.19
KIM	.05	-.08	.01	-.14
KIH	.14	-.01	-.11	-.22
KAL	-.10	.00	.00	-.01
KAH	-.29	-.17	-.05	-.07
Total score	.06	-.16	.02	-.24

Note. Total score is the sum of the component scores from the RC test. CC = Cognitive complexity. IC = Internal consistency.

Table 6

Simple correlation matrix for reading comprehension test, global structural characteristics, and student GPA in Experiment 1

Variable	2	3	4	5	6	7
1. GPA	-.08	.12	-.16	-.32	.04	-.10
2. RC Total Score (Pre-test)		.64**	.06	.04	-.16	-.07
3. RC Total Score (Post-test)			-.10	.02	-.12	-.24
4. Cognitive Complexity (Pre-test)				.35*	.15	.01
5. Cognitive Complexity (Post-test)					-.15	.21
6. Internal Consistency (Pre-test)						-.10
7. Internal Consistency (Post-test)						

Note. n = 38. RC = Reading comprehension.
* $p < .05$. ** $p < .01$.

There is an additional approach that can be pursued in order to help gain insight into what the iCmap paradigm is measuring, given it is not associated with reading ability, academic performance in introductory psychology or change in learner knowledge. By splitting participant responses into two groups, those concepts judged as related (i.e., assigned 1 during an iCmap trial) and those judged as unrelated (i.e., assigned -1), it is possible to aggregate the related and unrelated frequencies to produce two additional sets of group scores, analogous to the approach of Trochim (1989). By subtracting the observed frequencies after instruction from those frequencies obtained before instruction, a difference score is computed, which indicates the overall change in semantic judgments for a particular header concept. The difference scores obtained for the related judgments are presented in Figure 4 and are ordered from largest to smallest variance, from top to bottom.

There are two related patterns that emerge from this analysis. The first is that the variance in the difference scores is not homogeneous across the header concepts. Instead, what can be seen in Figure 4 is that the aggregated frequencies for some header concepts were consistently more variable than others. For example, "spontaneous recovery" had the highest variance of all the header concepts (SD =

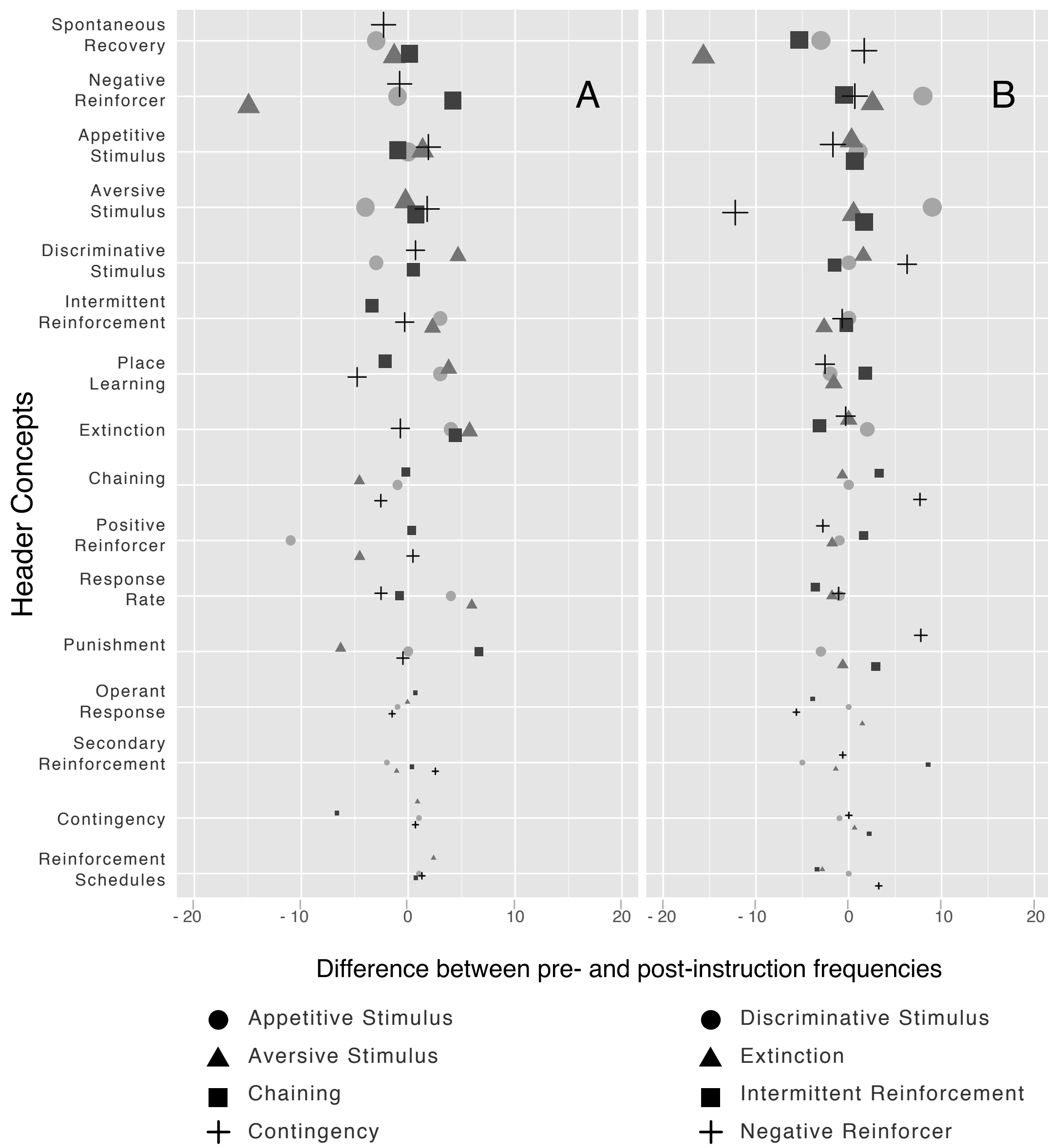

Figure 4

Difference in related categorization frequencies (n = 38) between pre- and post-instruction in Experiment 1. Header concepts listed from highest to lowest variance with symbol size indicating amount of variance.

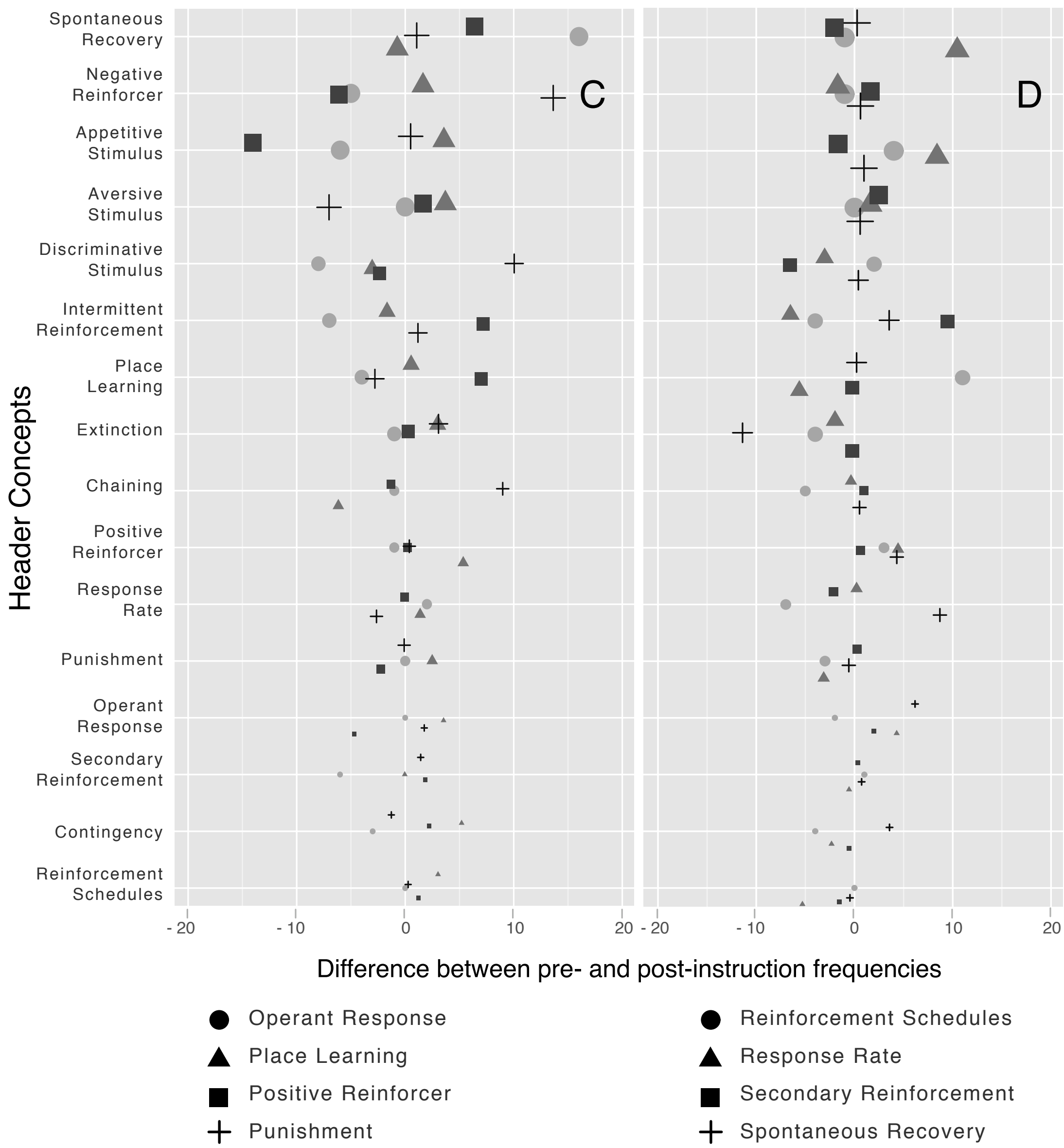

Figure 4
(cont)

7.05), while “reinforcement schedules” had the lowest (SD = 2.48).

The second pattern that emerged was a funnel shape owing to the difference between pre- and post-instruction relatedness scores; the greater the difference, the farther from center the concept pair was plotted. Importantly, some concepts were judged with greater frequency after instruction as “related”. For example, the concept “operant response” is rated as related to “spontaneous recovery” more frequently after instruction, while other concepts, such as “negative reinforcer” are rated as “related” to “aversive stimulus” less frequently after instruction. The header concept “spontaneous recovery” had a large sums of squares (SS = 746) compared to “reinforcement schedules” (SS = 92) at the other end of the spectrum, which observed very little change between pre- and post- instruction scores.

Discussion

In Experiment 1, a group of undergraduates were asked to categorize a set of concepts according to whether they were related or unrelated. In each trial, the participants were required to select two concepts as being related to a header concept, and two additional concepts as being unrelated to the header concept. The aims were to replicate the work of Aidman and Egan (1998) and Aidman and Ward (2002), and to improve the experimental design to assess whether learning would have an affect on the dependent variables, cognitive complexity and internal consistency. By collecting a set of proximity matrices before and after receiving instruction for the operational definitions of the concepts, it was possible to assess the validity of the dependent variables by first comparing the two sets of scores, and then assessing the amount of association with a reading comprehension co-variate, and finally assessing the association with each student’s final grade in the introductory psychology course where the concepts were taught.

The results indicate that neither CC or IC are sensitive to change due to learning, despite there being obvious changes in the underlying meanings of the concepts held by the students. Moreover, both CC and IC failed to correlate with the reading comprehension test, and the final grade. These results mirror the findings by Aidman and Egan (1998), and provide direction for several proposed refinements intended to improve the definitions of the dependent variables, and the design of the iCmap procedure.

The primary hypotheses under investigation in the present experiment were that CC and IC would both increase as a result of learning under the assumption that as students learn the definitions of the concepts, they would be able to identify more unique attributes, and be able to identify the correct pairings more easily. Paired sample t-tests of both variables indicated that there was no significant difference in the scores between the pre- and post-instruction. These results provide further evidence that the variables defined by Burmistrov and Shmeliov (1992) are not sensitive to conceptual structure per se, but rather merely describe properties of the scores within the proximity matrix. That is to say, the definition of CC to describe the number of unique dimensions within the proximity matrix does not appear to explain any of the variability in concept relatedness judgments, nor does it share any information with reading comprehension or the student's final grade.

The same can be said about IC given a lack of significant correlations with itself, the reading comprehension test and the final grade. One difference between the original work by Aidman and Egan (1998) and the present work was a failure to locate an association between IC and the final grade. One reason this could have happened was a change in the task design from three forced choices to four. Aidman and Egan did not provide any justification about why they chose one similarity and two contrastive judgments in their experiment, and because the algorithms prepared by Burmistrov and Shmeliov (1992) allow for a variable number of judgments per trial, moving to "two and two" (i.e., two related and two unrelated) per trial appeared to make the task more intuitive for the participants. Hence it is possible that the asymmetrical task design of Aidman and Egan favoured individuals with stronger underlying skills in concept categorization, whereas the modified design in the present experiment was easier to perform for all achievement levels. Further analysis is required to determine whether the number of concepts to select per trial has an effect of the structural properties of the proximity matrix.

The failure to locate significant differences necessitated a post hoc investigation of the data where the individual proximity matrices were partitioned into two subsets, one of just the scores indicating relatedness and the other into scores of non relatedness, under the assumption that formal instruction would provide accurate definitions, and explain both the operational and functional relationships characteristics of the concepts. The post hoc analysis revealed that the variability between the pre- and post-instruction related frequencies of the concept headers was not homogeneous. Instead what emerged was that some concepts had considerably more variability than others indicating that some of them had conceptual structures that were more complex and hence more difficult to categorize.

By inspecting which concepts held greater variability, a possible explanation for these results lies with the strategy adopted by participants during the pre-instruction phase of the experiment. Those concepts whose definitions were interpretable from the constituent terms, referred to as *semantically transparent*, could be reasonably guessed, and did not require meaningful learning to adequately understand or categorize. For these semantically transparent concepts, there was less variability between the pre- and post-instruction scores. On the other hand, those concepts whose definitions were not recoverable from the literal definitions of the constituent terms, referred to as *semantically opaque*, meaningful learning was required in order to correctly relate them. Thus, during the pre-instruction phase, it was likely that many of the categorizations of the semantically opaque concepts were simply guessed. In light of the fact that the task forced participants to make four choices whether they knew the definitions or not, randomly pairing concepts from trial to trial would appear to be the easiest strategy to adopt, with the consequence that any real effects would cancel each other out.

Although guessing is a reasonable explanation for the results of the pre-instruction phase, one would expect the post-instruction scores to move away from guessing towards applying the verbal definitions during the mapping process. However, the variance and central tendency of both CC and IC remained unchanged after instruction suggesting that despite the theoretical specification by Burmistrov and Shmeliov (1992) for the scores to range from zero to one, the information contained within the proximity matrices is so limited that there is little difference between them when they are reduced to a single value. It seems useful to consider ways in which more information can be added to the proximity matrix while still maintaining faithful to the overall iCmap procedure.

The results from Experiment 1 provide clues about the nature of the structural variables and the task design of the iCmap procedure, and where modifications might be fruitfully applied. By considering a few critical refinements to the task design, the iCmap procedure can be altered sufficiently to produce a proximity matrix that more accurately represents the underlying structures of the concepts presented during each trial. Several refinements are proposed in lieu of the results.

The first refinement focuses on the categorical assignment of concepts to the related and unrelated categories. The discussion presented in Chapter 1 revealed that concepts have a graded rather than a discrete structure, where people experience concepts as being more or less like the prototypical member. From that perspective, the iCmap procedure forces participants to collapse the network of mental models and their particular intensional relations along a singular dimension so that they can assert whether it is related to the header concept or not. When we consider the fact that both dependent variables are affected by the same state of affairs as described by Allen (2006) where considerably different concept maps can yield similar scores. In addition to experimental conditions that promoted guessing, it is not surprising that the expected results failed to be revealed.

The original choice to use categorical assignment came from the scoring procedure developed by Trochim (1989), but it is entirely possible to modify the task so that participants are able to rate the degree of relatedness by implementing a tool similar to a Likert-scale where an interval scale can be assigned to the judgments. Changing the measurement scale from nominal to interval for each concept pair is warranted statistically because the change would increase the variability contained in each score. The change is warranted theoretically because it would permit participants to express their sense of the graded relationship between the two concepts being rated.

Another way to add more information to the proximity matrix would be to require participants to rate all the concepts in each trial, not just the extreme ones. This would have the effect of filling each cell in the proximity matrix with relatedness information, which is substantially more than in Experiment 1 where only four out of the total 15 concepts were rated.

The second refinement that is motivated by the results of the present experiment is that some effort must be made towards developing a protocol for screening and pre-sorting the population of potential concepts so that the final set can be shown to be controlled on one or more dimensions. For example, the shortest concept, "chaining" is composed of eight letters, compared to the longest concept, "Intermittent reinforcement", which has 24 letters. Although these concepts well represented the domain of knowledge under investigation, the amount of variability their linguistic and structural characteristics contributed to the proximity matrices is unknown; therefore, in order to collect the cleanest scores possible, some attempt must be made to reduce the kinds of words that appear in each iCmap trial.

Arguably, the most problematic linguistic dimension left uncontrolled in Experiment 1 was an experimental list composed of both verbs and nouns, which according to Komatsu (1992) are ontologically distinct, and therefore encoded and processed in qualitatively different ways. This is rendered more salient within the mental model approach where mental models have structure that corresponds to the perceived or conceived structure of the state of affairs (Johnson-Laird et al., 1984) they encode. That is to say, the intensional information, the exemplars, the central tendencies and especially the background information between nouns and verbs must be considerably different. Some control must be enforced on the stimuli to bring the concepts more inline with each other.

The final refinement focuses on the limits of working memory and the effect that the number of words during an iCmap trial has on the network of mental models activated. Aidman and Egan (1998) did not base the number of concepts on any theoretical account of the limits of working memory, rather they set eight as the arbitrary lower threshold for how many concepts should be mapped together. Given that working memory has functional limits it is very likely that eight concepts might be too many for some participants to efficiently maintain and process. If the iCmap task can be altered to assess performance for fewer as well as more concepts, the added control might serve to illuminate response strategies, and even reveal effects washed out in the noise incurred from the original iCmap procedure. The discussion now turns to Chapter 3 where these modifications are applied and analyzed yielding a new procedure, called *progressive concept mapping*.

Chapter 3

From Models to Propositions

The results of Experiment 1 suggest that the iCmap task may be useful for assessing the conceptual associations among the words presented during each trial, a claim supported by the observation that the change in relatedness scores lacked homogeneity. However, the results also strongly indicated that the utility of the iCmap procedure is severely limited, primarily because the specifications and interpretations of the dependent measures yield no explanatory or predictive power in explaining the behavior underlying the observed scores or how those scores relate to conceptual structure.

In the second experiment, the amount of control imposed on the task design was increased by addressing several related issues. The first issue concerns the stimuli that were selected for the mapping procedure. In Experiment 1, the concepts were selected to represent a pre-defined domain of knowledge, resulting in a concept list that contained a mixture of word types including monomorphemic nouns and verbs, and compound words. In Experiment 2 that variability was constrained by selecting a single word type, while simultaneously filtering potential candidates according to several lexical characteristics (e.g., orthographic frequency). The increased control on the stimuli increases the likelihood that the processing demands placed on working memory during each iCmap trial will be consistent from trial to trial, thereby decreasing the amount of unexplained variance contained within each score.

The second issue concerns the iCmap scoring procedure and the measurements that are assigned to each rating. In Experiment 1, participants made only four ratings in each trial, a constraint that produced an extremely sparse similarity matrix. In Experiment 2 participants were asked to assign ratings to all the words presented in each trial. Collecting more pair-wise ratings increases the number of non-zero scalars that contribute to the final calculations used to obtain both CC and IC, which was expected to increase the range of observed values for both CC and IC.

From the mental model perspective, these added theoretical and methodological controls apply across the range of processing from the propositional codes that underlie and support the mental models that are activated during each trial to the models themselves. The lexical level factors controlled in the present experiment apply most directly to the propositional codes because features like word length and orthographic frequency are processed prior to conscious awareness and are not immediately salient as factors that contribute to perceived lexical meaning.

Shifting focus from models to propositions changes the perspective about how to conceive of the cognitive operations engaged during each iCmap task. The benefit that is gained, however, is tremendous because it establishes a common ground between research concerning conceptual structure and research investigating the psycholinguistic characteristics of language.

In the following section the kinds of words that should be used during the iCmap procedure are discussed first, followed by several theoretical accounts concerning the distinction between concrete and abstract nouns. The discussion then addresses the limits of working memory in language related tasks, and how the iCmap procedure can be modified so that information about working memory load can be used to interpret both CC and IC. The final topic that is presented highlights the need to establish a meaningful null hypothesis and what modifications can be made so that Experiment 2 can use the null hypotheses when evaluating the performance of both CC and IC.

Word Type Selection

In Chapter 2, it was mentioned that nouns and verbs are distinct because they map onto ontologically distinct aspects of the environment. Medin and Lynch (2000, p.125) further characterize nouns as "referring to clusters of correlated properties that create chunks of perceptual experience" whereas verbs focus on "relations among entities involving such things as causal relations, activity, or change of state." (p.125). Although either would be suitable within Experiment 2, nouns were selected because they tend to have more stable representations due to the clusters of correlated properties that they share.

The choice to use only nouns still leaves open a wide array of potential subclasses that can be used to partition the population of nouns into more refined classes such as abstract and concrete, basic level versus subordinate or superordinate level, or artifacts and natural kinds. Abstract and concrete nouns were chosen because they have been well-studied and there is a large body of work showing that they are processed in different ways.

According to Crutch and Warrington (2005), concrete and abstract nouns have representational systems that have qualitatively different properties. The observation that concrete nouns are better retained than abstract nouns, referred to as the *concreteness effect*, provides further evidence that these systems are qualitatively different (Ruiz-Vargas, Cuevas, & Marschark, 1996). There is also evidence that nouns and verbs have distinct neural representations (see Binder et al., 2005, for recent evidence and a review).

Ruiz-Vargas et al. (1996) discuss three different possible sources of the concreteness effect. The first possibility is an *elaborative processing* account that explains the concreteness effect in terms of the likelihood that concrete words will arouse more visual characteristics, and hence the propositional codes will be more strongly elaborated during subsequent processing. The second possibility suggests that there is differential enhancement of the associative networks instantiated between concrete and abstract nouns, which can be either based on the strength of the associations or the number of associations among concept nodes. The final possibility hypothesizes that concrete words have two different codes, with concrete words being underlain by both an *image-based* and a *verbal-based* code, and abstract words being underlain by a predominantly verbal-based code.

An example of such a theory is Paivio's (1991) Dual Code (DC) theory. DC theory states that when imagery is available (concrete words), recall is highly integrative even when verbal associations are weak, whereas when imagery is unavailable (abstract words), integration requires strong verbal associations. Ruiz-Vargas et al. (1996, p.47) point out that it is not the case that abstract words have no image-based propositions, but rather are "less likely to activate images or do so with greater difficulty because of their less direct access to the image-based system."

A different explanation for the concreteness effect comes from the work of Schwanenflugel and Stowe (1989), who proposed the Context Availability (CA) theory. Under the CA theory, there is only one representational system that serves both abstract and concrete nouns. The concreteness effect in tasks such as timed comprehension, naming and meaningfulness judgments is explained as being due to the the ease with which concrete nouns can access a network of relevant prior knowledge compared to abstract nouns. This explains why the concreteness effect, which is more pronounced when abstract words are presented in isolation, can be ameliorated by providing a richer stimulus context through the addition of a sentence or paragraph.

Schwanenflugel and Stowe (1989, p.117) argue that "abstract words are comprehended more slowly when presented in isolation because the reader is experiencing difficulty retrieving the relevant prior knowledge." The reason for this effect, according to Samson and Pillon (2004, p.253) is that "concrete words have greater contextual associations in semantic memory than abstract words." In support of this position, Crutch and Warrington (2005, p.623) claim that abstract words "may be acquired in the context of language without any direct perceptual input", which places importance on the linguistic context within which the abstract concepts are learned.

The CA theory is useful to consider in the present context because the set of concepts presented during each iCmap trial form a meaningful, albeit a limited, linguistic context. In Experiment 1, the linguistic context was highly informative because the concepts were selected to represent a particular domain of knowledge. The intention was that participants should use information from the context to shape their categorization ratings because the concepts shared not only definitional information, but also background and functional information.

In Experiment 2, the linguistic context was stripped of as much shared background information as possible so that each categorization judgment would place emphasis on the intentional information held by the concepts being rated. The hypothesis was that abstract words, because they rely on predominantly verbal-based associations, will produce larger complexity scores since participants need to access more intentional information in order to assign relatedness scores. The opposite trend will emerge for the consistency scores. Concrete nouns have salient image-based characteristics that will facilitate highly integrative processing. This should promote greater consistency in their ratings.

Working Memory

Working memory (Baddeley, 1986) can be thought of as the mental system responsible for holding and manipulating information during a variety of cognitive tasks. Working memory is implicated in a wide range of cognitive tasks including reading comprehension, skill learning, and complex problem solving (Schmiedek, Hildebrandt, Lovden, Wilhelm, & Lindenberger, 2009). It is implicated in language comprehension because different types of information must be integrated together into a coherent whole during language processing (Wagner & Gunter, 2004). One measure of working memory, reading span, captures the residual storage of working memory as well as the efficiency (Daneman & Carpenter, 1980). According to Dixon, LeFevre, and Twilley (1988) reading span, unlike traditional measures of short-term memory capacity, successfully predicts performance in reading comprehension tests.

Individuals with a low reading span can hold fewer items in working memory compared to individuals with a high reading span. Gunter, Wagner, and Friederici (2003) showed that individuals with a low span were not able to effectively inhibit irrelevant information, whereas high-span subjects were. One of the implications of the work by Wagner and Gunter (2004) and Gunter et al. is that low- and high-span readers use lexicon and contextual factors differently. Interestingly, Dixon et al. argue that the evidence suggests that the structural limits of working memory are less important than the efficiency of the processes operating for successful reading.

Working memory is very limited. As a result, people's performance declines rapidly with an increase in memory demand (i.e., the number of independent items held in working memory) in a wide range of experimental tasks (Oberauer & Kliegl, 2006). One model proposed to explain the limits of working memory, the *interference model*, assumes that items in working memory interfere with each other through interactions of their features (Oberauer and Kliegl). Oberauer and Kliegl point out that "when there are n items in working memory at the same time, each item suffers interference from n - 1 other items." (p.607). They go on to argue that the traditional "magical number" approach of viewing the capacity of working memory does not explain the "accelerated decline of asymptotic accuracy with increasing memory demand, together with a slowing of processing speed."

One of the challenges that the iCmap paradigm ought to address is the inherent difficulty in activating and maintaining a set of concepts in working memory. Where Aidman and Egan (1998) arbitrarily selected eight concepts to be presented in each iCmap trial, the design of Experiment 2 adds a second factor that manipulates the demand on working memory by gradually increasing the number of concepts so that the effect of memory demand on performance can be more thoroughly understood. The hypothesis is that there will be linear trends between the number of concepts to map and both complexity and consistency.

In the case of complexity, there will be a positive linear trend. More concepts will yield higher complexity scores. In contrast, consistency scores will have a negative linear trend because the demand on working memory will interfere with participants ability to hold the concepts and the relevant intensional information necessary to make concept rating. It is also predicted that number of words will interact with word type because concrete nouns will tend to have considerably more feature overlap, compared to abstract nouns. This should result in lower complexity scores for concrete nouns due to the interference of the activated features. Conversely, consistency will be weaker for abstract nouns because participants still must activate more intensional information in order to rate them.

Establishing a Null Hypothesis

Before turning to Experiment 2, there is one further topic to address that relates to the scoring procedure and usefulness of the iCmap procedure. Both initial iCmap investigations assumed that the task and scoring procedure faithfully described conceptual structure. However, when constructing a measuring device it is important to verify that the measurements reflect the absence of the phenomenon being measured when it is absent. The iCmap procedure forces participants to create associations among concepts because Aidman and Egan (1998) assumed that the concepts represented a domain of knowledge, and, therefore contained explicit associative relationships due to their verbal definitions, which each iCmap trial measured.

Theoretically, the iCmap task should only assign scores to meaningful conceptual associations and should assign a score of ‘zero’ in the absence of conceptual relations. There are two reasons why this is not the case. Firstly, participants were required to make four categorizations per trial under the assumption that the concept list does contain meaningful intentional information concerning their mutual associations. Regardless of participants’ inter-individual differences in the knowledge of the particular concepts, there is no way for a participant to indicate that there are no meaningful associations within a particular trial. Consequently, there is no possibility to create a similarity matrix that represents “no associations”, and hence there is no null hypothesis.

The second reason that a null hypothesis is difficult to formulate is because concepts are so deeply inter-connected. For all intents and purposes, it is likely that no set of concepts could be created for which participants would fail to find at least one characteristic that would allow them to apply some degree of structure on the similarity matrix. The absence of a null hypothesis makes it impossible to determine whether the conceptual structure imposed by the concept list has an effect on either complexity or consistency.

Because of these considerations, one of the modifications made to the task and scoring procedure of Experiment 2 was to make it possible for a participant to generate a similarity matrix with no associations, even though it is very unlikely that such a similarity matrix would ever be produced. This was made possible by setting one of the values they can assign to “no association”.

By combining the ability to create a null mapping with a concept list randomly generated to have no explicit associations, it becomes possible to collect a set of scores that closely resembles the null hypothesis. In actuality it is more likely that the scores collected under this perspective represent a baseline (‘low association’ rather than ‘no association’) condition which can be used to compare the similarity matrices from different concept lists that do or do not contain explicit conceptual relations.

In Experiment 2 another factor is added to the experimental design that uses two different concept lists for each type of noun. One list was completely random, and the other was structured to contain explicit relations among at least two of the concepts in the list.

The hypothesis under investigation was to assess whether a third factor, explicit semantic associations in the *random* condition would be lower for complexity than in the *explicit* condition. This was expected because participants would have little or no intentional information shared amongst the concepts in the random condition. In the explicit association condition participants would identify the imposed structure above and beyond the verbal definitions of the concepts, and incorporate the contextual information into their ratings. The same pattern was expected to emerge in the consistency rating, because the presence of explicit semantic information would facilitate more consistent ratings compared to the random condition.

Experiment 2

The goal of Experiment 2 was to adapt the iCmap procedure to account for a set of theoretical and methodological refinements which were argued to provide increased control over the scores assigned by participants to each pair-wise rating. The scores provide the ability to re-assess the usefulness of the structural variables, which are used to describe characteristics of the mental models maintained and manipulated in working memory during each trial. If the scores are sensitive to the particular factors manipulated within the current experiment, then a more meaningful interpretation of the structural variables may be developed, which goes beyond the work of Aidman and Egan (1998).

Two critical changes were made to the iCmap procedure, which warrants a new name for the task: progressive concept mapping (proCmap). The first change added two additional concepts to each block beginning with three concepts (i.e., one header concept, and two concepts to rate) in the first block and ending with nine concepts in the fourth block. The second modification changed the categorical assignment adopted by Aidman and Egan (1998) into a continuous score. The end result of developing the proCmap procedure is a tool that can be used by researchers and educators interested in further understanding conceptual structure and the process of meaningful learning.

Method

Experiment 2 assessed the proCmap procedure in a 2 (abstract vs. concrete) x 2 (random vs. structured) x 3 (5, 7, or 9 words per trial) mixed factors design with a repeated measure on the last factor. The first factor was generated by selecting words according to imageability and concreteness ratings from Coltheart (1981). The second factor was created by randomly generating a list of words for the random condition, and by replacing one word in the random list so that it would have an explicit semantic association with other words in the structured condition. As explained above, the third factor was created by adding two additional words to each block, beginning with three words in the practice block and ending with nine. This resulted in four blocks to which each participant contributed scores, with only the last three blocks contributing to subsequent analyses. The same reading comprehension co-variate as used in Experiment 1 was used in Experiment 2.

Participants

One hundred and twenty University of Alberta undergraduates (78 female and 42 male; M_{Age} = 19.0, SD = 1.73) volunteered to participate in the experiment. If the participant did not consent then an alternative activity was provided of equal educational purposes. Regardless of whether the participant completed the experiment or activity, he or she received partial course credit.

Materials

Reading Comprehension Test. Participants were seated in separate rooms at a computer terminal with the Mac OS X 10.4.0+ operating system and a 15" LCD computer monitor. Input was collected from a standard keyboard and mouse. The RC test used in Experiment 1 was again used in Experiment 2. The only change to the RC test was to add a unique identifier to each test statement so that reliability estimates could be calculated, which corrected a problem encountered in Experiment 1.

proCmap. Though the proCmap task is different from iCmap task used in Experiment 1, it retains the essence of the iCmap task. In Experiment 1, participants were required to drag concepts into graphical "bins" (related/unrelated) which assigned one of two values [1,-1] to the concept. In Experiment 2, each concept was displayed above a slider that enabled participants to slide their rating along a scale. There were n sliders drawn on the computer screen, where n is the number of concepts to map in the current block (see Figure 5). Each slider was 250 pixels wide, which when presented on monitors with a resolution of (1024 X 768) resulted in a scale that was 93 mm in length. The slider itself occupied 5 percent of the total line length.

Each slider was divided into two equal segments where the center point was used to represent a lack of knowledge or a guess about the relationship between the concept and the header, and is assigned zero. The range of scores remained limited between 1 and -1, as in the original specification of the iCmap procedure, however each rating assigned a decimal value equal to its proportional distance from the centre point. The proCmap task preserves the directionality of the values implemented Aidman and Egan (1998), but allows participants to provide graded information about the relatedness of the concepts relative to the header.

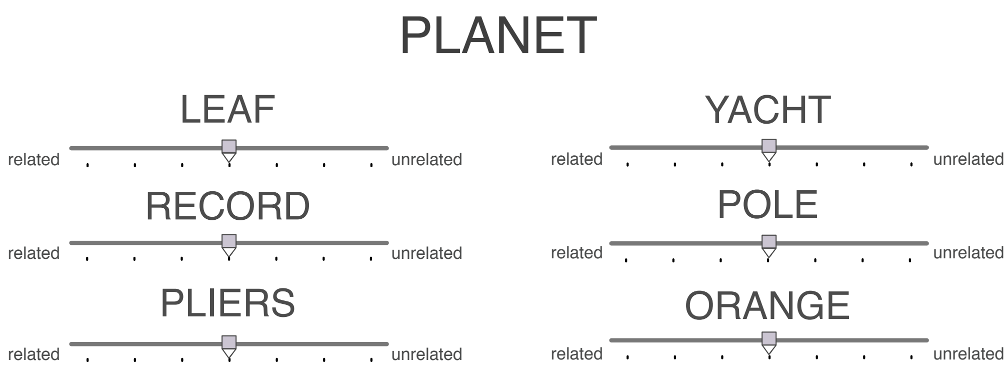

Figure 5
Graphical illustration of the progressive concept mapping task developed for Experiment 2. Each trial began in the null hypothesis state depicted in this illustration.

The stimuli lists that were presented in each block of the proCmap procedure were constructed with a more rigorous procedure than in Experiment 1. Following the work of Binder et al. (2005) two variables were selected for matching the abstract and concrete stimuli lists for Experiment 2: orthographic length and orthographic frequency. Orthographic length is the number of letters in a word. Orthographic frequency (oFreq) is the number of occurrences of a word per million words of written text. Imageability ratings taken from Coltheart (1981) were used to partition nouns into highly imageable (concrete) and low imageability nouns (abstract).

Because there were two word type conditions, two semantic conditions and four blocks of trials within each condition, 16 stimuli lists were created by matching words according to the variables above. First, words were partitioned in abstract and concrete according to their imageability ratings. Next, a three-word stimuli list was created by randomly selecting words from the available set of concrete (or abstract) nouns. The structured three-word stimuli list was identical to the random list because the first block was used as a practice block, thus no modifications were necessary. The five-word stimuli list was created by adding two more randomly selected words to the three-word list.

The lists differed in the seven- and nine-word conditions. The procedure to randomly select two additional words was again used to construct the seven-word list. However, for the structured list, one of the words was exchanged to contain an explicit association with other words in the list. For example, in the seven-word random concrete list, the word *lemon* was selected. To create the seven-word structured concrete list, lemon was replaced with the word *orange* because it shared a physical feature with the word *planet*, namely both are round or spherical.

Finally, the nine-word random list was created in the same fashion as the seven-word list, with two additional words being selected and added to the list. The nine-word structured list was created in the same way that the seven-word structured list was created, only this time the word *napkin* was exchanged for *autumn* because one of the colors of autumn is orange, and both autumn and orange share perceptual features with sunsets.

Constructing the abstract stimuli lists followed the same procedure as outlined above. The full set of 16 stimuli lists are located in Appendix B. The descriptive statistics for the four variables used to match words in Experiment 2 are presented in Table 7.

Procedure

Reading Comprehension Test. Participants were greeted and given a consent form to read and sign. The consent form described the nature of the experiment and provided detailed information about their rights while participating in the experiment. If the participant did not consent to participate they would have been explicitly assured that they would suffer no academic sanction and given an alternate activity to

Table 7
Characteristics of concepts by condition used in Experiment 2

		No. of letters[a]		Word Frequency[a]		Imageability[b]		Concreteness[b]	
Word type	Semantics	*M*	SD	*M*	SD	*M*	SD	*M*	SD
Concrete	Random	5.3	.87	17.8	20.2	601.7	22.9	580.0	39.8
	Structured	5.4	.88	22.8	20.2	605.4	21.7	561.0	65.5
Abstract	Random	5.3	.87	27.7	20.8	337.8	37.8	309.9	46.7
	Structured	5.4	.88	40.6	32.0	330.6	35.1	289.9	42.6

Note. n = 120. No. = Number. *M* = Mean. SD = Standard deviation.
[a] See Baayen et al., 1995. Word frequency based on occurrences per million.
[b] See Coltheart, 1981.

complete. No participant requested the alternative activity. The remainder of the procedure for the RC test was the same as in Experiment 1.

proCmap. Participants remained in the same room after completing the RC test. Each participant began the proCmap session with both text-based and verbal instructions indicating that they would be required to read a list of words located in the center of the computer display, and that they would have to rate those words according to their meaning by dragging the slider along the scale located directly beneath each word. Participants were specifically instructed to treat each end of the scale to represent that they *absolutely certain* that the word was or wasn't related to the header while the center of the scale was to be used to indicate that they did not know what the relationship was between the two concepts was and therefore represented *no knowledge*.

Each trial initialized all the sliders to the center of the scale, which placed each trial into the null hypothesis state (see Figure 5). When all words had been assigned ratings, participants indicated that they had completed the trial by pressing a button on the screen, which initiated the next trial. This process repeated until each word in the list had taken its turn as the header. At the beginning of each subsequent block, the participants were told that more words were going to be added, but the task remained the same.

Before initiating the first block, participants were informed that it was a practice block so they could familiarize themselves with task, and that if they had any questions to contact a research assistant. Finally, participants were informed that there would be four blocks to complete, and that they would be provided with instructions at the start of each block so they didn't have to memorize the instructions.

Data analysis

The resulting similarity matrices were then analyzed to compute the structural variables.

Cognitive Complexity. The data analysis for CC followed the same procedure that was used in Experiment 1, except the cell values were assigned the value that the slider was moved to, normalized to be between -1 and 1.

Cognitive Consistency. The calculation of Internal Consistency was based on comparing the sign of the scalar values from the rows with the scalar values from the columns from the similarity matrix. When two scalars had the same sign, they were conceptually placed in the same pile, and if they had different signs, they were not. As a result of moving to continuous values along with the intention of preserving the magnitude of the cell values, Equation 2 developed by Burmistrov and Schmeliov (1992) was not used. Instead the Pearson correlation (r) among the scalars was used to assess the degree of consistency between the rows and columns of the similarity matrix. To prevent confusion with the results obtained in Experiment 1, a new variable was created, *cognitive consistency* (CI), which is the correlation of the row and column scalars in each block.

Results

Reading Comprehension

The dependent variable for the RC test and the subscales was accuracy (i.e., number correct). The descriptive statistics for the RC test are presented in Table 8. None of the component processes suffered from ceiling or floor effects, and there was a wide range of scores for each. The simple correlation matrix for the RC test total score and subscales along with the correlations from Experiment 1 (in parentheses) are shown in Table 9. Performance on the RC test in the present experiment was similar to the performance observed in Experiment 1.

Table 8
Descriptive statistics for reading comprehension test in Experiment 2

Component	*M*	SD	Range
Text-based memory	62.9	11.47	(27−82)
Text-based inference	25.0	5.54	(13−35)
Knowledge integration low	20.3	3.12	(11−24)
Knowledge integration med	29.2	4.58	(15−36)
Knowledge integration high	26.0	5.32	(16−35)
Knowledge access low	32.4	2.97	(20−36)
Knowledge access high	21.6	2.24	(12−24)
Total score	217.3	30.22	(144−265)

Note. Total score is the sum of the components. n = 120; *M* = Mean. SD = Standard deviation. See Table 2 for component maximums.

A reliability analysis, which was not possible in Experiment 1, was carried out. The result of the analysis was an extremely high Cronbach's index (α = .96). This suggests that the RC test is a reliable measure of reading comprehension and a useful co-variate. The test was, however, lengthy, requiring roughly 20 minutes to complete. The purpose of including the RC test as a co-variate was to determine if there were chance differences among the groups on reading ability that might have been introduced by randomly assigning participants to the experimental conditions.

Table 9
Simple correlation matrix for reading comprehension test in Experiment 2

Component	2	3	4	5	6	7
1. TBM	.84** (.83)	.70** (.44)	.80** (.47)	.80** (.63)	.43** (.21)	.40** (.22)
2. TBI		.61** (.36)	.72** (.47)	.71** (.61)	.51** (.05)	.43** (.21)
3. KIL			.76** (.57)	.68** (.68)	.67** (.64)	.54** (.39)
4. KIM				.82** (.71)	.57** (.42)	.52** (.31)
5. KIH					.44** (.45)	.44** (.30)
6. KAL						.66** (.41)
7. KAH						

Note. n = 120. Experiment 1 correlations are in parentheses.
* $p < .05$. ** $p < .01$.

Submitting the scores to an mixed factors ANOVA revealed that the co-variate was not different among the groups for either variable: ($F_{CC} < 1$; $F_{CI}(1,115) = 3.54$, $p > .05$). Therefore subsequent analyses proceeded without adjusting the condition means.

proCmap

The two dependent measures derived from each similarity matrix were Cognitive Complexity, and Cognitive Consistency. In order to provide a transparent discussion of the results, the distribution of each variable will be discussed first, and then the descriptive statistics for both variables will be presented.

Cognitive complexity performed similarly to the pattern of results in Experiment 1, as can be seen in Figure 6. In all four conditions, CC scores tended to cluster around .75, with very few scores falling below .5. There were several outliers identified. Given the lack of research to-date with these variables there was no formal procedure to adjudicate whether to remove them or not, so they were retained in subsequent analyses. Interestingly, there were fewer outliers in the Structured conditions than the Random conditions.

Cognitive consistency followed suit with observations again converging around

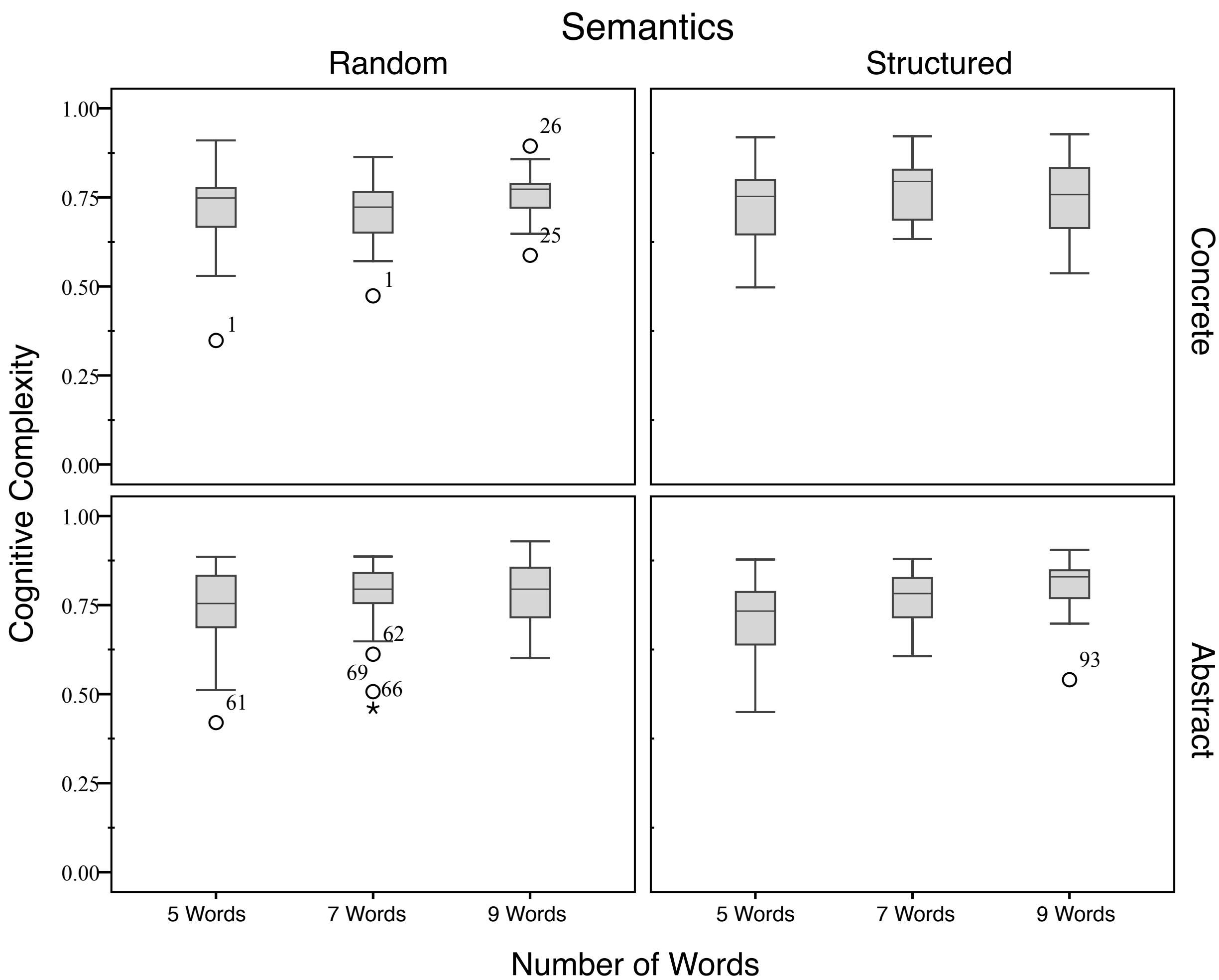

Figure 6

Boxplots of Cognitive Complexity by condition in Experiment 2

.75 (see Figure 7). However there was one notable exception, namely there were values that approached zero, which means that some participants applied no systematicity to their ratings. More peculiarly, there were more outliers for CI that appeared in the Structured condition than the Random condition. Again, outliers were retained in subsequent analyses.

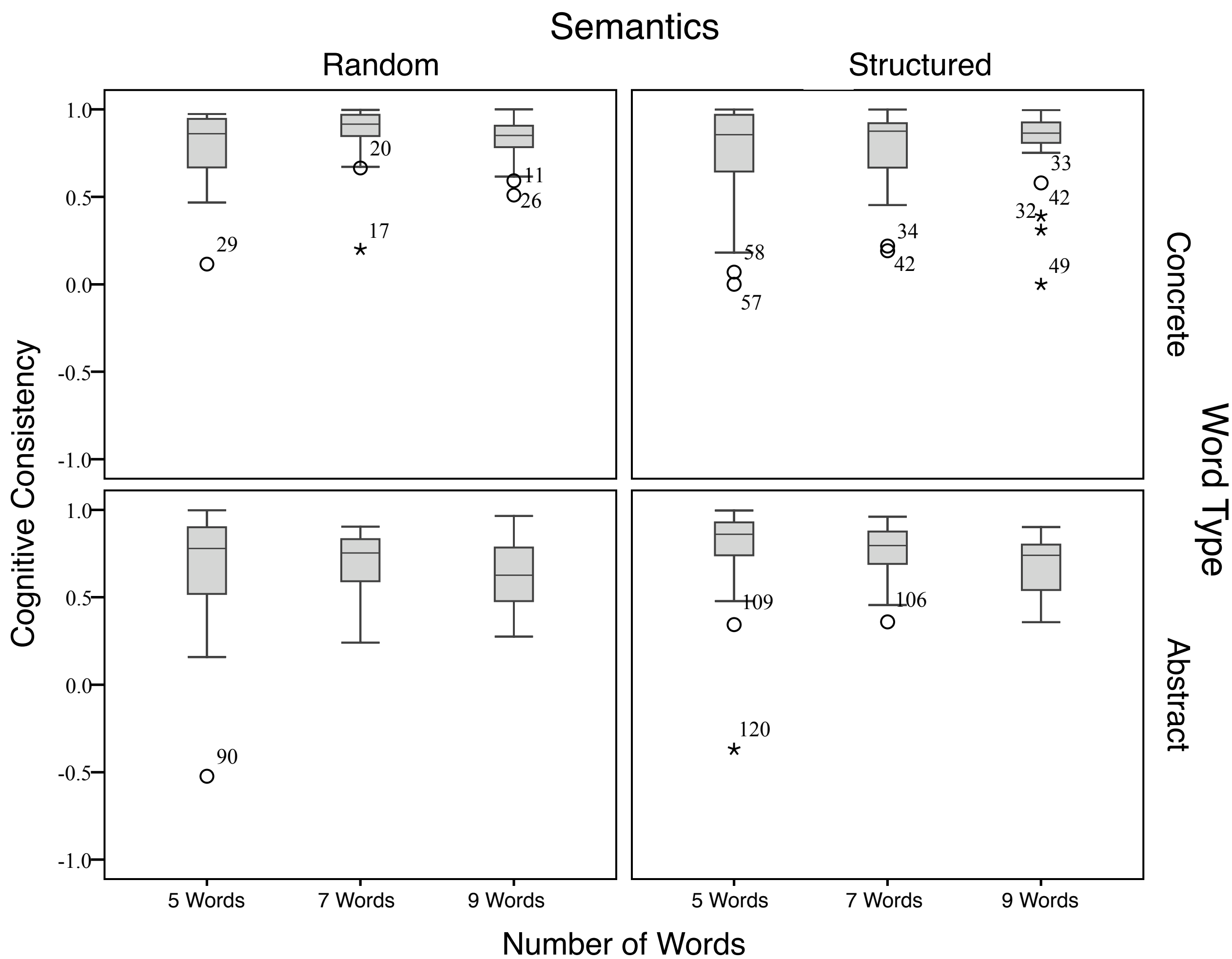

Figure 7
Boxplots of Cognitive Consistency by condition in Experiment 2

The marginal means for Word Type are presented in Table 10. There appeared to be agreement with the prediction that abstract words would tend to have higher complexity ratings compared to concrete words ($M_{Abstract}$ = .77, $M_{Concrete}$ = .74; F(1,115) = 4.48, p = .036). The prediction that concrete words would have higher consistency ratings was also supported ($M_{Concrete}$ = .81, $M_{Abstract}$ = .70; F(1,115) = 15.78, $p < .001$).

Table 10
Marginal Means for Word Type on CC and CI in Experiment 2

	Cognitive Complexity		Cognitive Consistency	
	M	SD	*M*	SD
Concrete	.74	.10	.81	.21
Abstract	.77	.10	.70	.22

Note. M = Mean. SD = Standard Deviation.

The marginal means for Number of Words are presented in Table 11. In the case of CC, it was predicted that there would be a positive linear relationship, which was suggested by the means (M_{5word} = .72, M_{7word} = .75, M_{9word} = .77) but no supported statistically (F(1,115) = 1.88, $p > .05$). Additionally, it was predicted that more words would impede cognitive consistency ratings resulting in a negative linear trend. This hypothesis was not supported by the data either (M_{5word} = .75, M_{7word} = .79, M_{9word} = .73; F < 1).

Table 11
Marginal Means for Number of Words on CC and CI in Experiment 2

	5 Words		7 Words		9 Words	
	M	SD	*M*	SD	*M*	SD
Cognitive Complexity	.72	.11	.75	.09	.77	.09
Cognitive Consistency	.75	.27	.78	.18	.73	.19

Note. M = Mean. SD = Standard Deviation.

However, there was an interaction between Number of Words and Word Type on CI (F(1,115) = 4.05, $p = .047$). This interaction is depicted in Figure 8, which reveals that Concrete words are progressively rated more consistently, while Abstract words are progressively rated with less consistency.

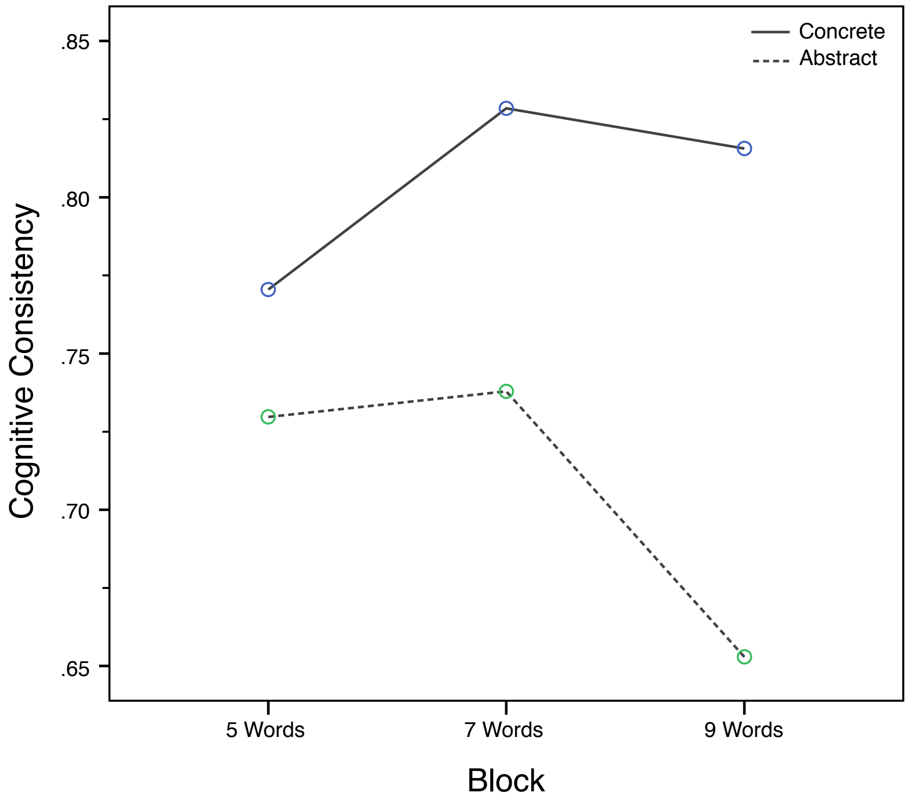

Figure 8
Interaction between Number of Words and Word Type on CI in Experiment 2

The descriptive statistics for cognitive complexity on the Semantics factor at each level of Word Type are presented in Table 12. First, the cell means for the 5-word conditions were predicted to be equal because the stimuli were identical. The data reveals that was the case (M_{Random} = .72, $M_{Structured}$ = .72). The cell means for the 7-word condition were predicted to be smaller in the random condition, compared to the structured condition. This prediction was marginally supported in the means from 7-word, concrete condition (M_{Random} = .70, $M_{Structured}$ = .77), but not in the 7-word abstract condition (M_{Random} = .77, $M_{Structured}$ = .77) or in either 9-word condition (M_{Random} = .76, $M_{Structured}$ = .75; M_{Random} = .78, $M_{Structured}$ = .81). These patterns were not significant, however, revealed by a non-significant three-way interaction for the linear trend of Number of Words by Word Type by Semantics ($F(1,115) = 1.41$, $p > .05$), and a non-significant linear trend of Number of Words by Semantics ($F < 1$). The three-way omnibus interaction between Word Type, Semantics, and Number of Words on CC that was significant ($F(2,115) = 3.45$, $p = .033$) indicating that some other trend in the cell means was present.

Table 12
Descriptive statistics for Cognitive Complexity in Experiment 2

		5 Words			7 Words			9 Words		
Word Type	Semantics	*M*	SD	Range	*M*	SD	Range	*M*	SD	Range
Concrete	Random	.72	.11	(.35−.91)	.70	.08	(.47−.86)	.76	.06	(.59−.89)
	Structured	.72	.12	(.50−.92)	.77	.08	(.63−.92)	.75	.10	(.54−.93)
Abstract	Random	.74	.12	(.42−.89)	.77	.11	(.46−.89)	.78	.09	(.60−.93)
	Structured	.72	.11	(.45−.88)	.77	.07	(.61−.88)	.81	.07	(.54−.91)

Note. M = Mean. SD = Standard Deviation.

The descriptive statistics for cognitive consistency on the Semantics factor at each level of Word Type are presented in Table 13. There is one notable trend in the cell means, which is that in the 5-Word Abstract conditions, two participants produced negative consistency ratings. This is intriguing because it means that these participants were changing the ratings they assigned between the header and the concepts from ‘related’ to ‘unrelated’ between trials. This could indicate guessing, not responding honestly, or that over the course of the task they identified new information that weighed in on their beliefs about the relationship among the concepts.

Again, the cell means for the 5-Word conditions were predicted to be equal because the stimuli were identical. However, the means reveals that the Random concrete words were rated with slightly more consistently than the Structured concrete words ($M_{Random} = .79$, $M_{Structured} = .75$). The cell means for the 7-word condition were predicted to be smaller in the Random condition, compared to the Structured condition, a trend which appeared to emerge in the 7-word Abstract condition

Table 13

Descriptive statistics for Cognitive Consistency in Experiment 2

		5 Words			7 Words			9 Words		
Word Type	Semantics	*M*	SD	Range	*M*	SD	Range	*M*	SD	Range
Concrete	Random	.79	.20	(.12−.97)	.86	.16	(.20−1.0)	.82	.12	(.51−1.0)
	Structured	.75	.28	(.00−1.0)	.79	.22	(.19−1.0)	.81	.22	(.00−1.0)
Abstract	Random	.68	.31	(−.52−1.0)	.70	.17	(.24−.90)	.63	.19	(.28−.97)
	Structured	.77	.30	(−.37−1.0)	.77	.15	(.36−.96)	.68	.15	(.36−.90)

Note. *M* = Mean. SD = Standard Deviation.

(M_{Random} = .70, $M_{Structured}$ = .77). However the opposite trend emerged in the 7-word Concrete condition (M_{Random} = .86, $M_{Structured}$ = .79). In regards to the 9-word condition, there appeared to be no difference of Semantics in the Concrete condition (M_{Random} = .82, $M_{Structured}$ = .81), but a small advantage in the Structured Abstract condition over the Random Abstract condition (M_{Random} = .63, $M_{Structured}$ = .68). The above patterns were supported by a significant linear trend of Number of Words by Word Type interaction ($F(1,115) = 4.05$, $p = .047$), but no linear trend for the Number of Words by Semantics interaction, nor the three-way linear trend interaction.

Post Hoc Analyses

There are two tacks to approach further exploration of the results from Experiment 2. The first addressed further statistical patterns and the second pursues an inspection of the graphical plots produced from the similarity matrices. In the first case, there is only one further analysis that is needed to describe the patterns found in the cell means. It was predicted that there would be linear trends as a result of progressively adding more words to each trial. What emerged however were two quadratic trends.

The first quadratic trend occurs between Number of Words and Semantics on

CC ($F(1,115) = 7.90$, $p < .01$), but the three-way quadratic is also significant ($F(1,115) = 8.48$, $p < .01$) thus the quadratic is best interpreted from the three-way (see Figure 9).

This reflects the fact that, in the Concrete condition, the Structured concepts gained complexity between the 5- and 7-word blocks, but then decreased in complexity, whereas in the Random condition the complexity measure first decreased and then increased to roughly the same level as in the concrete condition. There was no quadratic trend in the Abstract condition, in which complexity progressively increased. These trends are also presented in Figure 10.

Cluster plots. The final analyses are qualitative and follow Aidman and Egan (1998) and Aidman and Ward (2002) who inspected the graphical representations of similarity generated from each similarity matrix. For the sake of brevity and clarity, only the cluster diagrams are discussed because they are somewhat easier to interpret. To begin this discussion, the first cluster plot that is discussed is the null hypothesis (see Figure 11).

Based on the task design used in experiment 2, there are three ways a participant could produce the null mapping: 1) rate all words as completely related, 2) rate all words as completely unrelated, and 3) make no ratings at all. The effect in all three scenarios is to produce a similarity matrix with no variability in the ratings. The null mapping is useful as a baseline to compare the following cluster plots because it is now immediately possible to see the implicit structure that participants are expressing in their ratings. The discussion will address the cluster plots from the 9-Word conditions by presenting those maps that were the best and poorest performing participants on each DV.

In the Concrete Random condition (see Figure 12), the top two cluster plots depict the implicit structures created by the participants with the highest scores in

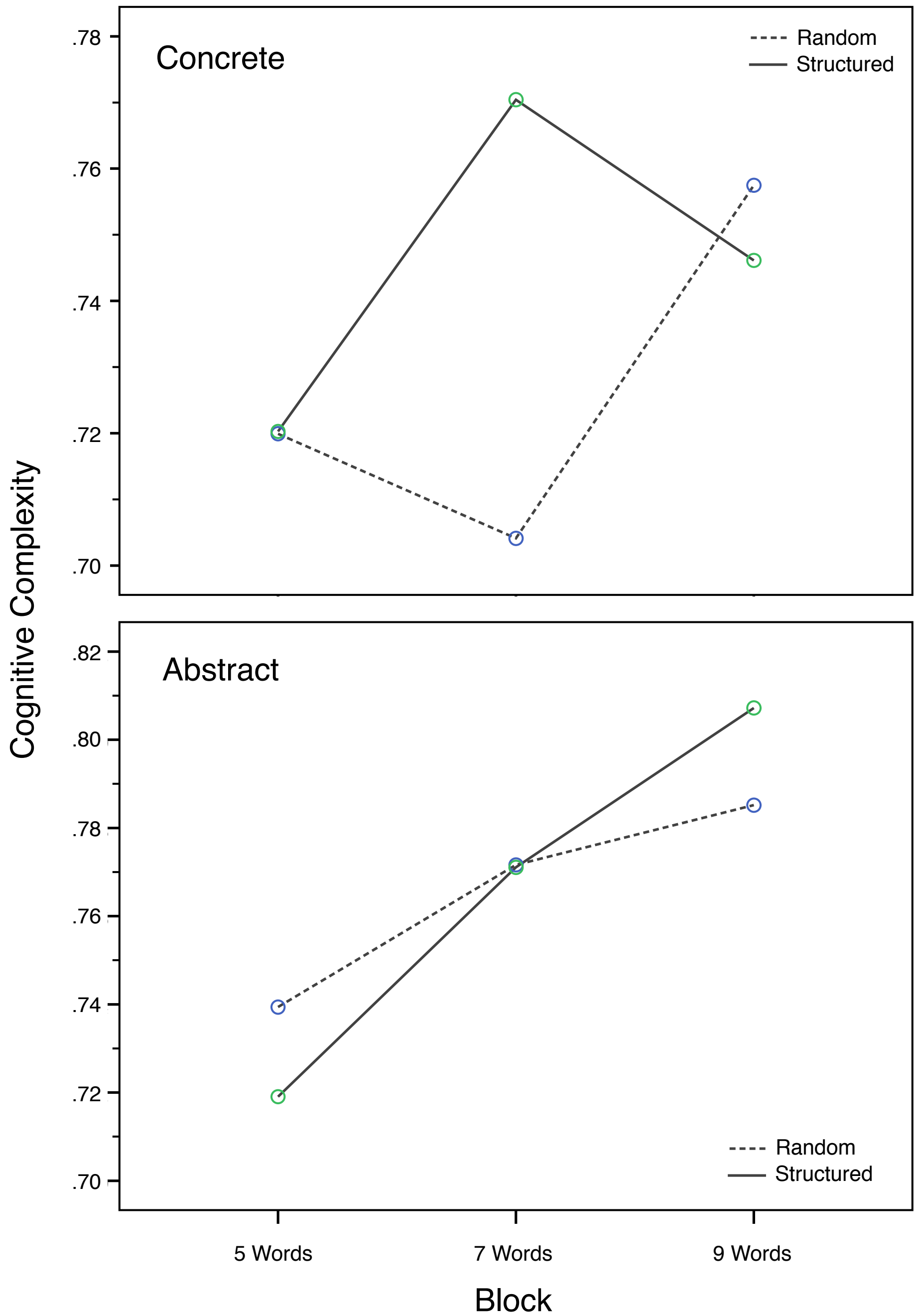

Figure 9
Three-way Interaction between Word Type, Semantics and Number of Words on CC in Experiment 2

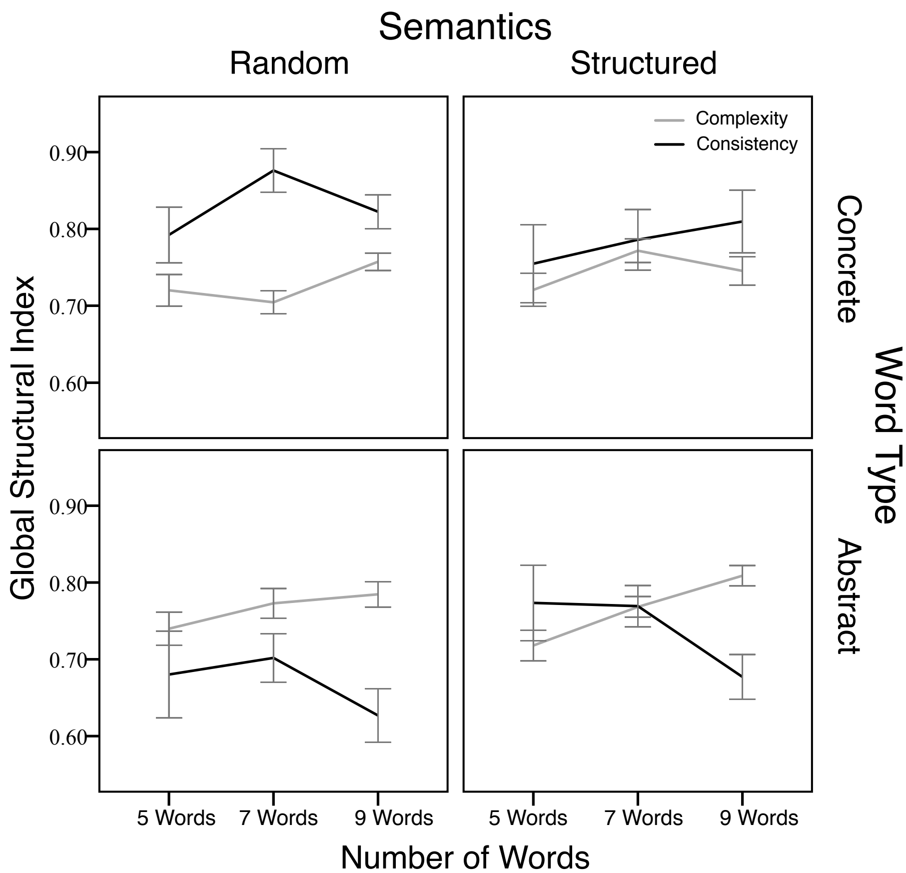

Figure 10
Effect of number of words on CC and CI for each condition in Experiment 2

Figure 11
Hierarchical cluster plot for the null hypothesis.

CC and CI. In the case of CC, the participant scored .89, a relatively high score and produced a very deeply structured cluster plot. Inspecting the plot revealed that concepts were organized around the well established distinction between natural nouns versus artifact nouns. Clearly, naturally occurring objects clustered near the bottom and progressively becoming less related based on size and shape, until finally one section of the plot contained just man-made objects.

Importantly, when interpreting cluster plots, it is important to consider the branching structure as well the depth of the linkages.

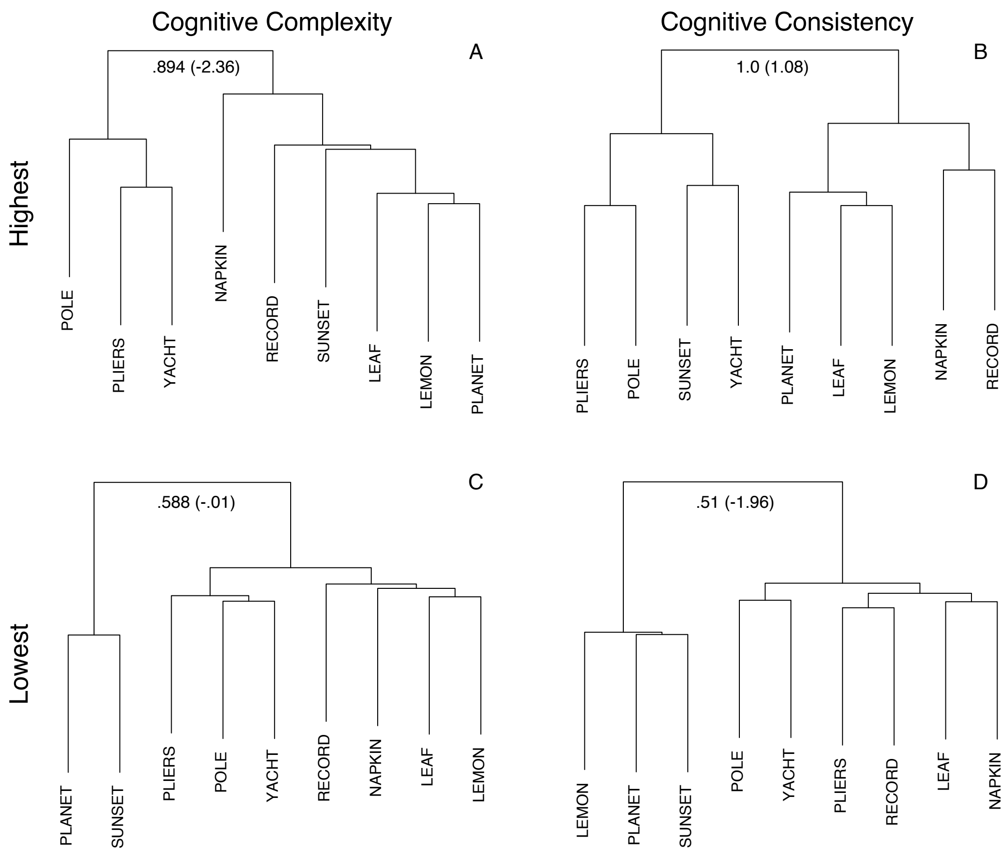

Figure 12
Hierarchical cluster plots for the highest and lowest performing participants in the Concrete Random group. The score for each measure is listed in the diagram along with the standardized reading comprehension score in parentheses.

Even though the words *pliers* and *yacht* are depicted at the same depth as *leaf*, because they are located on different branches, they are very dissimilar. Interestingly, the next plot in Figure 12 denotes the highest consistency cluster plot, which appears to be more evenly distributed. The words appear to cluster around three primary properties: 1) natural vs. nominal, 2) size, 3) orthography. Moreover, the CI score of 1.0 indicates that the participant was perfectly consistent in assigning the ratings while maintaing the three properties. To help visualize how the proCmap task created these scores, two example trials from participants scoring the highest CC and highest CI in their condition are presented in Figures 13 and 14.

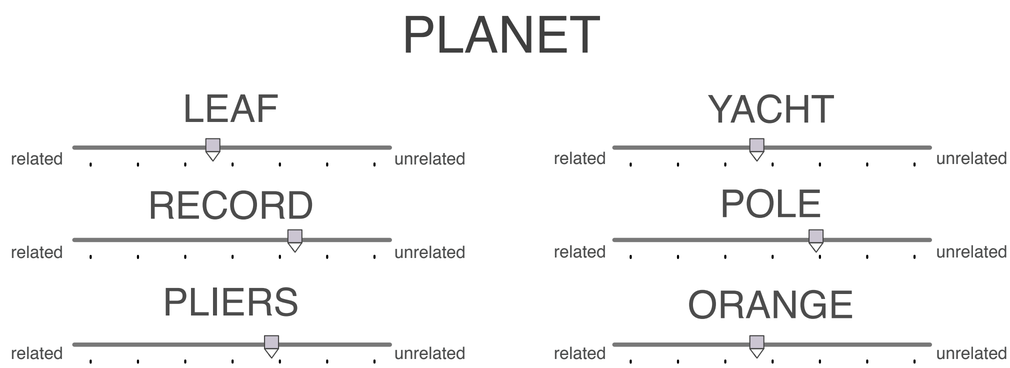

Figure 13
Example trial from from the participant who scored the highest CC in the concrete systematic group

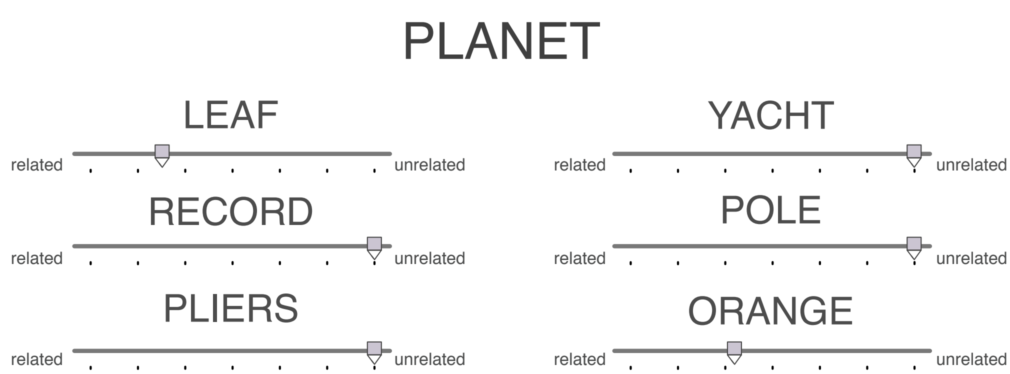

Figure 14
Example trial from from the participant who scored the highest CI in the concrete systematic group

In contrast, the last two cluster plots in Figure 12 reveal the poorest performing participants and the first feature that stands out is that the properties around which the words are classified are based heavily on the perceptual characteristics of the words. That being said, *planet* and *sunset* were consistently grouped close together indicating a salient imageable characteristic was accessed when rating these words.

Inferring the other characteristics leads the discussion into the terrain of idiosyncratic meanings that are not entirely useful for understanding the results of Experiment 2. Many convincing theories could be proposed to account for these particular cluster representations, therefore using the cluster plots to infer general characteristics of conceptual structure is not particularly helpful.

Inspecting the next set of cluster plots (see Figure 15) for the Structured Concrete words does reveal a significant difference in the structure. Recall that the Structured list swapped two words from the Random list (i.e., *orange* for *yellow*, and *autumn* for *napkin*) with the intention of bringing *orange*, *sun*, *autumn*, and *planet* into coherence. Notably, only the participant who scored 'zero' on consistency didn't capture the above cluster of words. In every other case, there is an obvious partitioning of words into those experimentally intended to be similar and those that were random.

Comparing the cluster plots from Figure 12 and 15, reveals that the proCmap procedure is sensitive to explicit associative information above and beyond the verbal definitions of the words.

Turning to the cluster plots for the abstract words (see Figures 16 & 17) a similar picture emerges. When words were selected to bring coherence to the list, participants successfully located that coherence and captured it in their ratings during the proCmap task. Those participants that used superficial characteristics tended to produce cluster maps organized around orthographic properties, perceptual properties, and weak verbal associations. For example, the participant who scored the lowest CI in the Structured Abstract condition judged the words *theory* and *guess* to be very related indicating their belief that theories are like guesses, which reveals

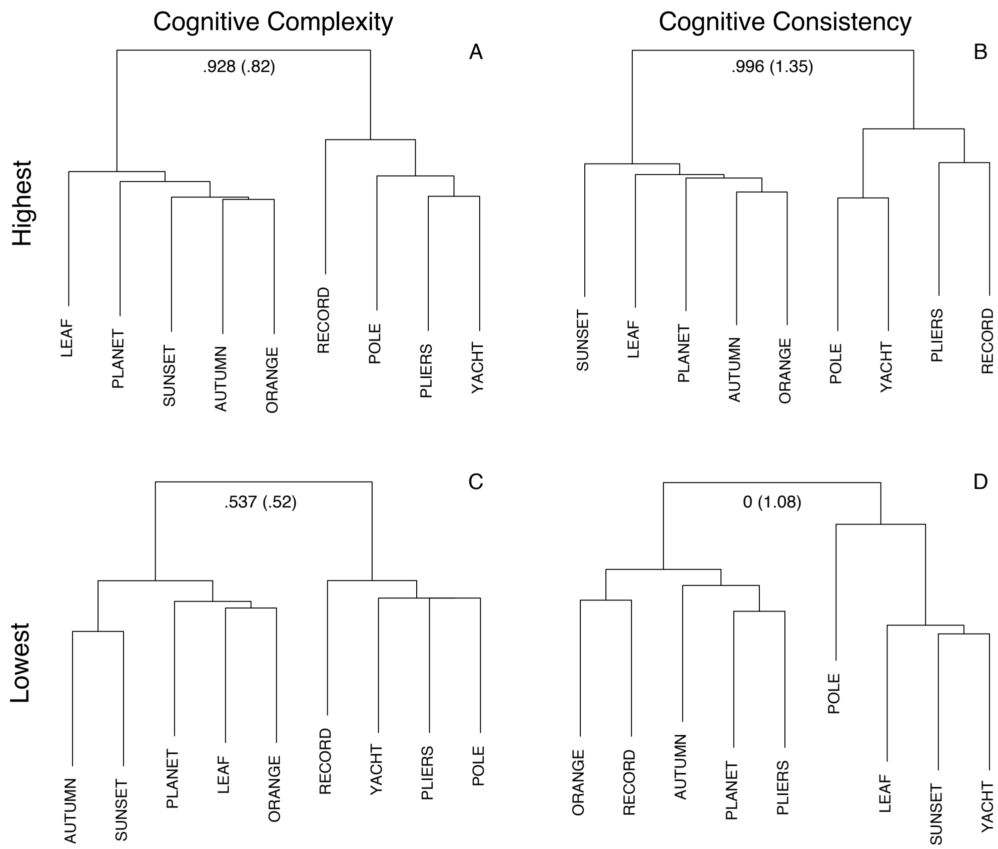

Figure 15
Hierarchical cluster plots for the highest and lowest performing participants in the Concrete Structured group. The score for each measure is listed in the diagram along with the standardized reading comprehension score in parentheses.

a potential misunderstanding of the meaning of the word theory. It would appear that the cluster plots do provide useful qualitative information that each participant could use to learn about the quality of the knowledge.

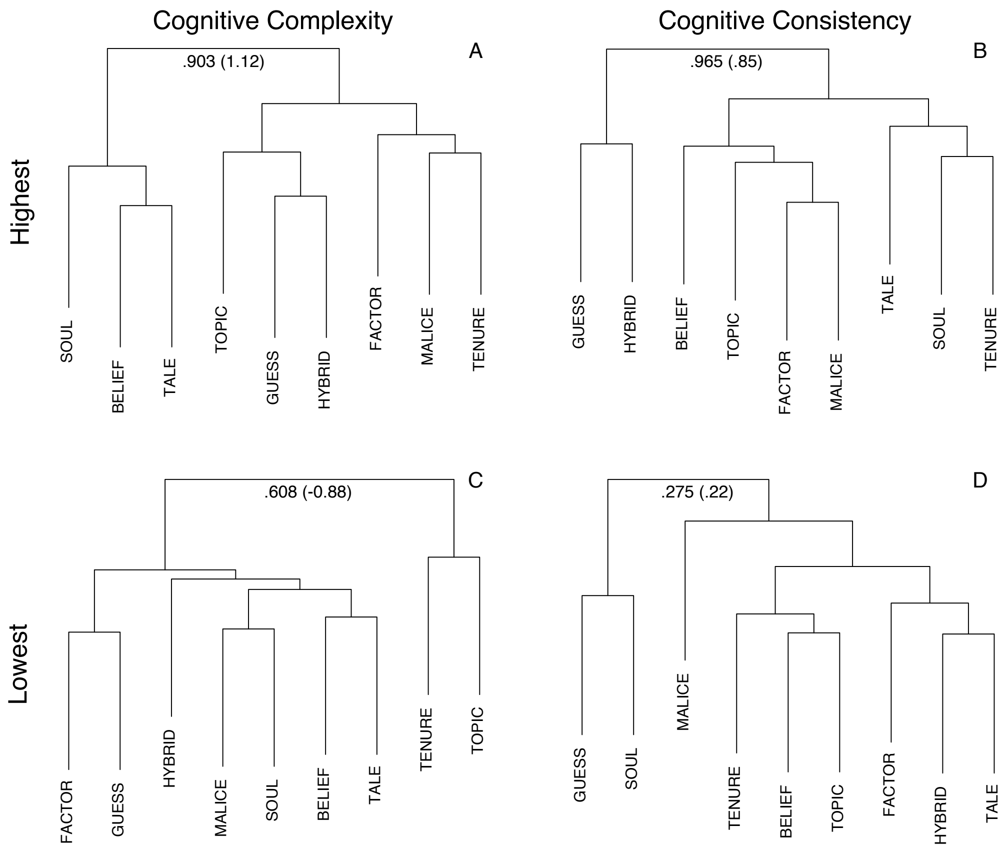

Figure 16
Hierarchical cluster plots for the highest and lowest performing participants in the Abstract Random group. The score for each measure is listed in the diagram along with the standardized reading comprehension score in parentheses.

Discussion

In this study, a group of undergraduates was asked to rate the degree of semantic relatedness among a list of words presented in a new task called progressive concept mapping. The aim was to assess the efficacy of the proCmap methodology at describing two useful, related characteristics of conceptual structure: the complexity of the conceptual information activated during each trial, and the consistency that

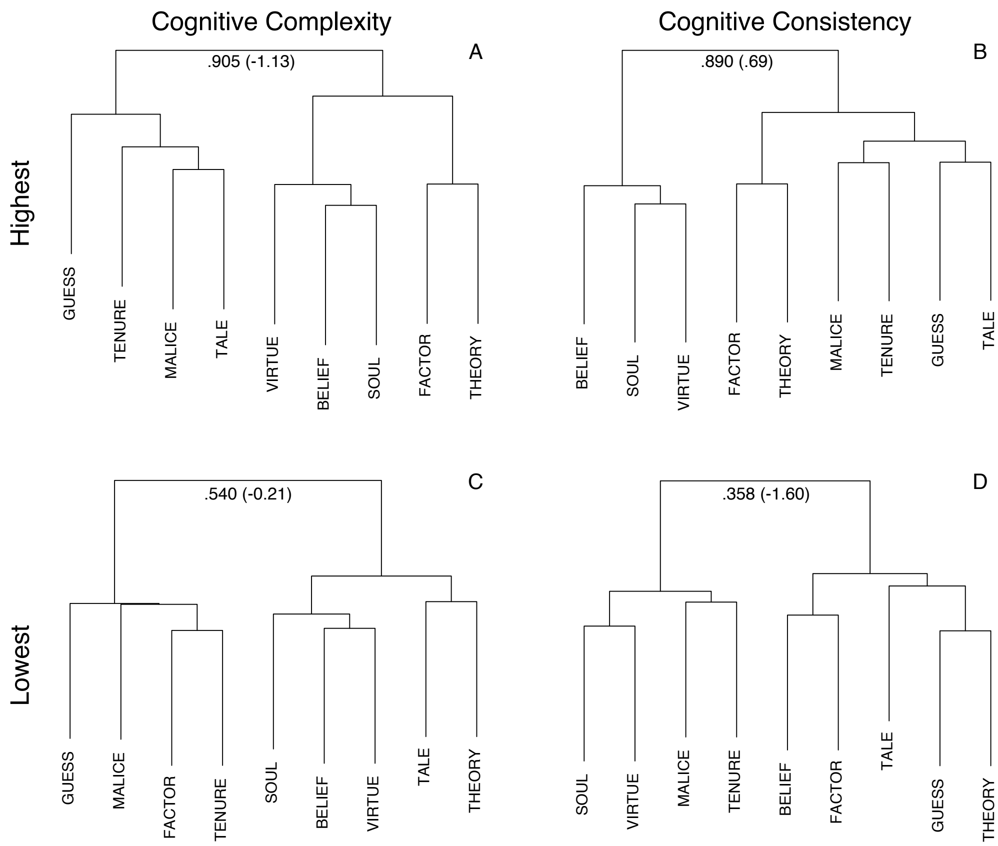

Figure 17
Hierarchical cluster plots for the highest and lowest performing participants in the Abstract Structured group. The score for each measure is listed in the diagram along with the standardized reading comprehension score in parentheses.

conceptual structure can be expressed in this particular computer-based task.

By assessing the performance of two variables, Cognitive Complexity, and Cognitive Consistency, across a range of factors, it was possible to answer a set of theoretically motivated questions capable of providing explanations concerning the nature of the scores collected for both dependent measures. Three factors were manipulated, word type, number of words, and semantic content in order to test the effect that each had on the ratings assigned by participants. In addition, the number of letters, and orthographic frequency of each word were used to match candidate items selected for each level of the above factors.

The first factor, word type, partitioned the population of potential words into nouns with high imageability (concrete), and those with low imageability (abstract). It was predicted that imageability would play a crucial role in the kinds of propositional codes activated between concrete and abstract nouns that would in turn have several functional benefits for concrete nouns (i.e., a concreteness effect) such as highly integrative processing, and the activation of highly salient, image-based propositional codes. Abstract nouns on the other hand would only have indirect access to image-based propositional codes and would rely on predominantly verbal-based propositional codes. The results provide converging evidence for the differential processing of concrete and abstract nouns, by supporting both predictions that concrete words would be rated more consistently than abstract words ($F(1,115) = 15.78$, $p < .01$) and with less complexity than abstract words ($F(1,115) = 4.47$, $p = 0.036$). These results are the first to demonstrate that the proCmap procedure is sensitive to the underlying representational structure of concepts. However, these results require further qualification due to the interactions they had with the remaining factors.

The next factor tested was number of words consisting of three levels: five words, seven words, and nine words. Each level placed greater demands on working memory in an attempt to see if the processing of concrete nouns relative abstract nouns would reveal two opposing trends. In the first case, it was predicted that a positive linear trend would emerge between the number of words and CC scores because a larger linguistic context would activate and process more intensional information in working memory. CI on the other hand was predicted to decrease with the number of words due to increased demands on working memory thereby reducing the available resources required to recall previous ratings. Neither of these predictions were supported.

However, what did emerge was an interaction between number of words and word type on CI ($F(1,115) = 4.05$, $p = .047$). This pattern indicates that concrete nouns facilitated consistent responding over the course of the experiment such that demand on working memory did not appear to have detrimental effects on the concept ratings. Abstract nouns appeared to be negatively impacted from the load on working memory in the nine word condition. It is difficult to adjudicate between the various theories attempting to account for the concreteness effect as to which best explains the concreteness results obtained here, because the experiment was only intended to assess the proCmap procedure. However, it appears that greater linguistic context did not facilitate consistent responding, thus it is more likely that the DC theory provides the simplest explanation. During the course of each trial, with all things being equal between the concrete and abstract conditions, the additional image-based propositional codes serve to anchor relatedness judgments more effectively than the predominantly verbal-based codes of abstract words.

Finally, the semantic factor established a baseline condition for conceptual relatedness scores compared to a more realistic condition wherein some of the words had explicit associations with increased complexity for the structured words as well as increased consistency. Neither the two-way linear interaction between the semantic factor and number of words nor the three-way linear interaction between all three factors was significant for CC. Performance of CI was the same. Post hoc analyses identified two quadratic trends as a two-interaction between semantics and number of words and a three-way interaction. These results indicate that the seven word condition tended to be the optimal number of words to compare, which more strongly supports Dixon et al.'s (1988) claim that the structural limits of working memory are less important than the efficiency of the processes. These results are less supportive of Oberauer and Kliegl's (2006) claims that working memory declines rapidly with an increase in memory demand because performance first increased from the 5- to the 7-word condition and then decreased in the 9-word condition.

There are several limitations in the present study. The first is that re-casting consistency as a correlation changes the index that CI is measured against. Thus even though the condition averages between CC and CI do tend to be in the same range of values, comparing them directly is not appropriate. It was done here for descriptive purposes to give a sense of how these variables behaved, but future research will need to revisit the two structural variables. In light of the results obtained in Experiment 2, it seems that CC is not a very useful variable for describing stable characteristics of conceptual structure, as such it should be left out of future research until it can be re-formulated and re-validated.

The second limitation of the present experiment was the specification of the semantics factor. The imposed coherence of the structured list was not based on any formal specification, and given that it was an arbitrary choice about which words would promote coherence, the effect of semantics must be interpreted with caution. Further research should explore ways in which conceptual coherence can be incrementally imposed on the word lists.

The third limitation was the specification of the number of words factor as a repeated measure. Although participants did rate the same initial three words in all four conditions, the five-, seven- and nine-word conditions were not the same as the three- word condition. A better description of this factor is required in order to truly assess the impact that working memory load has on conceptual rating scores.

Despite the above limitations and need for further empirical support for the results obtained here, the initial results are promising. It was argued that the refinements made to the iCmap task would yield a more robust task that was better able to measure conceptual structure. The results provide initial validation of that claim that the proCmap procedure is a useful tool. Most importantly is the observation that the kinds of words selected (concrete vs. abstract) to represent the domain of knowledge interact with the amount of overt associative information among the words. Thus, researchers and educators who use this kind of concept mapping or more traditional concept mapping techniques should pay close attention to the words they choose. That being said, in order for this procedure to really be effectively used within other experimental work on conceptual structure, serious attention must be given to the specification of the structural variables. Interestingly, the graphical representations that are output by this procedure continue to provide the most useful information about an individual‘s understanding of a particular domain of knowledge, which is exactly where Aidman and Ward (2002) ended their investigation of the iCmap paradigm. This can be interpreted as suggesting that the dynamic and context sensitive nature of concepts is better represented graphically then as a single value.

Future research into the proCmap procedure should consider re-framing the theoretical perspective used to situate conceptual structure. Although mental models provide a robust account of concepts, new work by Gabora et al. (2009) begins the work of redefining concepts within an ecological perspective. “[I]t is only when objects in the world have been conceptualized that they are charged with the potential to dynamically interact in myriad ways with the conceptions of other objects as well as with the goals, plans, schemas, desires, attitudes, fantasies, and so forth, that constitute human life." (p.95). Gabora et al. go on to state that "it is when stimuli in the world come to be understood in conceptual terms that they acquire the weblike structure and self-organizing dynamics characteristic of ecology." (p.95).

In sum, the proCmap procedure has demonstrated that it is a useful tool, and that it warrants further attention because it is sensitive to the well-established distinction between concrete and abstract words. This result provides converging evidence of the concreteness effect and that the proCmap procedure can detect this effect. Moreover, it opens the door to experiments that allow researchers to investigate conceptual structure in real-time, and under contexts that more meaningfully resemble the way in which concepts are used, in service of interacting with the world.

Chapter 4

Conclusion

In the present manuscript I set out to assess and validate the iCmap paradigm, specifically the measurements that the task collected, whether it was a useful tool for measuring conceptual structure, and ultimately whether it was capable of generating new research questions. The iCmap procedure was originally designed to help students learn how to use conventional concept mapping based on the observation that students benefit from seeing the relationships implicit in their knowledge. The benefit of both implicit and explicit concept mapping for the student is a meaningful learning experience where new information is appropriately associated with pre-existing knowledge structures.

The primary focus of the iCmap paradigm is on the contents of concepts, and the associative links between concepts because the task requires individuals to read a list of words, and to then classify those words into discrete categories based on their shared meanings. Because the iCmap task was designed to measure implicit conceptual structure, research into this paradigm necessarily overlaps with the work of other researchers in cognitive psychology and psycholinguistics who are investigating lexical and semantic access; however, there are important differences that had to be addressed.

The goal of the first chapter was to present a theoretical framework capable of describing conceptual structure so that the results of the experiments could be meaningfully interpreted. I argued that concepts should be viewed as mental models that encapsulate highly complex, and dynamic structures, which are context-dependent and richly inter-connected. Such models are also analogous to the states of affairs that they encode, suggesting, for example, that concrete nouns activate very different models from abstract nouns, which are very different from verbs and so on.

To evaluate the dependent measures generated by the iCmap paradigm, I designed an experiment to assess whether changes in conceptual structure due to learning were detectable by the iCmap task. The hypothesis that I was interested in assessing was simple. The iCmap procedure generates numerical scores that reflect characteristics of the underlying conceptual structure as well as the conceptual processing during each trial; therefore, formal training should increase the available information contained within concepts, and clarify the most relevant information necessary to make judgments about those concepts. More specifically, I asked whether formal training would increase both cognitive complexity, and internal consistency as indicated by larger scores in the post-instruction condition.

The results of Experiment 1 revealed that formal training did not have any effect on the distribution of scores that the iCmap procedure assigned to the classification judgments. The null results suggested that the tool was either not sensitive enough to detect the changes, or that the task design was not sufficiently rigorous to remove extraneous factors inflating the amount of noise contained within each classification judgment.

Despite the null results, the iCmap task still appeared to be an intuitive and sensible way of asking questions about the way in which people use concepts to interact with the world. Many important characteristics concerning language production and comprehension have emerged from researchers using the semantic decision task (a relatively simple task), which made the iCmap task appear useful to successfully generalize beyond a single pair-wise decision to a set of pair-wise decisions. In fact, the most useful information to be gleaned from the first experiment was that a change in conceptual structure had taken place, but it was only detectable by comparing the pre- and post-instruction scores for each of the header concepts (comparing all pair-wise judgments across each header concept), not by comparing the individual scores assigned to participants. The reason for this originates from the nature of the scoring procedure. A single trial produces a sparse, n-dimensional vector because only four pairings are required in each trial. In order to compare the performance of the header concepts across for the entire experiment, the trials for all the participants were aggregated together. Consequently, there was considerably more information pooled together when comparing the header concepts.

By pooling the scores together a pattern emerged that provided a key insight that proved to be useful for deepening our understanding of the semantics of the header concepts and how to build better stimuli, namely, that some concepts have definitions that are transparent whereas others have definitions that are opaque. However this pattern was not useful for classifying individuals into meaningful categories.

These results did reinforce the decision to adopt a pre-post design as the ideal way to investigate change in conceptual structure and whether the iCmap procedure was sensitive to changes in conceptual structure, that is to say the experimental design was sound. The additional control of within-subject variability and the ability to answer whether the iCmap procedure validly describes conceptual structure makes the pre-post design ideal. By having the students complete the task before instruction and then afterwards was the correct way to assess the dependent variables under their current specification.

The failure to find an association between the dependent variables and a reading comprehension test or academic achievement suggested that the iCmap task was not describing the complexity of the conceptual structures or the processes used to compare them. Certainly the task assigned numerical scores to the classification judgments, but the results called into question the validity of those scores.

In retrospect, had I thought carefully about why Aidman and Egan (1998) neglected to discuss the dependent variables in more detail, and their complete lack of explanation about the relationship between the scoring algorithm and conceptual structure I would have come to the conclusion that a problem existed. The results from Experiment 1 clearly followed the original investigations, and called into question the interpretation of the scores that were being assigned. While the iCmap task remained an intriguing way of looking at conceptual structure, the above problems presented an important challenge that I believed could be resolved. Something had to be done to improve that state of affairs so that the dependent variables could be used to partition individuals based on their performance in the task.

The purpose of the second experiment was to increase the amount of information collected from each participant, and to greatly improve the task design, which was renamed to proCmap - indicating that the number of words per trial was no longer fixed. I controlled a few structural and probabilistic features of the words so that differences in word length and frequency were minimized and contributed as little extra noise as possible. As well, I modified the way that numerical scores were assigned to each rating, opting for continuous instead of dichotomous values because research had shown that categories are graded, not discrete.

Additionally, I asked if the kinds of concepts (concrete or abstract) presented during each trial could facilitate processing, and whether the number of concepts increased demand on working memory. For the final modification I wanted to develop a null hypothesis, a situation where no judgments were made to act as a baseline to which observed classifications could be compared to. My choice to develop a null hypothesis differs from the original work by Aidman and Egan (1998) where they produced an expert map, and the participant maps were compared to the expert map. Because that there are many possible expert solutions, and many ways for participants to achieve an expert mapping, developing a null condition seemed more useful.

The behaviour of the dependent variables in the second experiment followed the first in terms of the range of scores that each group produced. This was not expected because I predicted that by adding more scores on each trial in conjunction with continuous values, that there would be more variability in the range of scores, potentially spanning the theoretical range of zero to one specified by Burmistrov and Shmeliov (1992). That was not the case. Thus the evidence suggests that there is a fundamental problem with the cognitive complexity scoring algorithm. That is to say, the task does not actually produce data that conforms to the claims made by the original authors.

The new variable, cognitive consistency, on the other hand did produce a wider range of scores, with a couple of outliers falling below zero. Cognitive consistency is simply the average correlation among the scalars in the dissimilarity matrix computed from the raw classification scores. The fact that the new consistency variable obtained values near zero indicates that, either some individuals guessed from trial to trial, or that their knowledge tended to be superficial and rote, which resulted in inconsistent classification scores. In order to produce an average correlation of zero, one would have to invert the classification ratings from trial to trial, which would be difficult to do because of the way that the task is designed. From this data we now have a way to establish a cut-off value for participants who obtain a low score that indicates the presence of guessing.

Nevertheless, the second experiment primarily confirmed the existence of the concreteness effect. On the surface this finding is not profound, but it does suggest that the proCmap task, a completely different task compared to semantic categorization, is capable of detecting the effect in a sample of university undergraduates. There were two noteworthy interactions that emerged. The first interaction was between Word Type, the number of words, and the semantic factor on cognitive complexity. I predicted that there would be a linear trend such that more words per trial would increase the complexity of the classification judgments. While that trend emerged for abstract nouns, concrete nouns revealed two quadratic trends: if the nouns were selected at random, individuals began by using less information, and then as more words were added, more information; if the words were chosen to fit a pattern, then individuals began using more information and then switched to using less information. The divergence between the two cannot be accounted for by either mental models or by dual-coding theory, and so leaves an open question for future research to address.

The second interaction emerged with cognitive consistency between the number of words and Word Type, which verified my prediction that, as more words were added to each trial, concrete nouns would have greater consistency. This prediction was based on the argument that highly imageable nouns are strongly integrative because the perceptual features of the non-verbal representations are very salient. Abstract nouns on the other hand only have verbal representations; therefore, more words should impede performance in terms of consistency, and this was observed.

The results suggested that the proCmap procedure was a valid way of describing aspects of the conceptual structure insofar as it was used within the task implemented in Experiment 2. One characteristic to emerge from the results of Experiment 2 was the importance of theoretically motivated decisions when designing and implementing experiments. By controlling the stimuli in Experiment 2 along several dimensions, the proCmap procedure was able to tease apart significant differences in performance across the four conditions. In contrast, modifying the number of judgments per trial, adopting continuous scores, and varying the number of words per trial produced very little additional information relative to the effects due to Word Type.

The results from Experiment 2 were helpful in highlighting the assumptions and oversights that were built into the experiment because they indicated that cognitive complexity was not a valid measure of conceptual structure, and that the scoring algorithm had to be completely re-worked before any new research could be carried out with the proCmap task. The failure to reveal linear trends with number of words was also problematic because the quadratic trends could not be explained. I might have anticipated the null effects based on the results from Experiment 1 given that it was a sound experiment that failed to reveal effects for cognitive complexity. The three-way, higher order interaction on cognitive complexity though interesting, does not appear to be a useful effect given that cognitive complexity requires a considerable overhaul.

Overall the experiments confirmed that there are problems with both the iCmap and the proCmap procedures. The most problematic is the nature of the algorithm that computes cognitive complexity. This algorithm is the mechanism that associates the classification judgments to numerical scores, the scores that are used to assess conceptual characteristics. Because the scores appear to be relatively insensitive to individual differences in conceptual structure, it does not seem appropriate to continue using this variable in any future research.

Changing the formulation and interpretation of the consistency score appeared to produce useful results. I would argue that cognitive consistency does describe an interesting characteristic of conceptual processing. Namely, that the conceptual structures activated and maintained in working memory are not simple geometric shapes that are arranged and rotated until they fit together. Rather structure and process are intimately linked, and that performance measures need to assess the intentional relationships from more than one direction to get a clear picture of what the concepts mean to the individual. Taken together, the experiments presented in Chapter's 2 and 3 were successful in providing the evidence necessary to definitively argue that cognitive complexity is not a useful variable, whereas cognitive consistency appears to be. Without the research carried out within this manuscript, it would have remained uncertain why the initial work by Aidman and Egan (1998) and Aidman and Ward (2002) turned out the way that it did. In that sense, the research in this thesis helped to clarify the terminology used to discuss concepts, reveal the diverse range of theories used to explain and understand conceptual structure, and shed light on a handful of the factors that contribute to conceptual processing.

Future Directions

The above discussion has identified several problems with both the iCmap and proCmap procedures. These problems are easily addressed and can be incorporated in another experiment. First and foremost, I recommend simplifying the task so that it uses a fixed number of words to present in a session, (i.e., revert back to the iCmap task structure). This will greatly reduce the statistical problems that emerged by adopting a variable number of words. Although I believe it is important to be inclusive and work towards bridging the gap between various domains within Psychology, those researchers who specialize in working memory should propose modifications based on theoretical and empirical evidence. Until then, focus on the iCmap paradigm as a tool to investigate the characteristics of concepts.

The second modification would synthesize the experimental design of Experiment 1 with the rigour of Experiment 2. Instead of selecting words based on their lexical and probabilistic features, I recommend that a set of non-words be constructed along with fictitious definitions. Creating fictitious definitions will ensure that all participants have the same baseline exposure to the meanings of the stimuli, and will help remove variability due to prior exposure. Additionally, the definitions can be created with different levels of semantic complexity. For example, one could adopt the strategy used by Hannon and Daneman (2001) where the number of features could be manipulated or the number of relations among the definitions. Designing the definitions based on theoretical dimensions would ensure that in the pre-instruction condition participant ratings would all be near or at baseline. Then there will be an instruction phase where the definitions for the words are presented. The definitions should be created so that there are levels of understanding implicit within them so that some of the stimuli are similar and others are not. In the post-instruction condition participants repeat the iCmap task. This design will ensure that all participants completing the pre-instruction condition will have as close to identical understanding as possible (i.e., no knowledge).

Finally, the scoring algorithm used to summarize the dissimilarity matrix should be addressed. The purpose of the scoring algorithm is to assign a single numerical score to a set of classifications as a way of describing the underlying structure within the dissimilarity matrix. The Q-technique developed by Stephenson (1953) should be used as an alternative way of interpreting the data, but other approaches should not be ruled out. The work of Allen (2006) may provide insight into methods for summarizing the data collected during the iCmap task.

The research has been instrumental in clarifying and furthering my understanding on the nature of concepts, and I hope that future research can refine my initial steps outlined here so that cognitive psychology and linguistics can build a unified theory of concepts that is capable of situating the vast array of experiences that concepts participate in with the practicalities of modern research.

References

Aidman, E., & Egan, G. (1998). Academic assessment through computerized concept mapping: Validating a method of implicit map reconstruction. *International journal of instructional media*, *25*(3), 277.

Aidman, E., & Ward, J. (2002). Implicit concept mapping: A computerized tool for knowledge assessment in undergraduate psychology. *Methods of Psychological Research Online*, *7*(3).

Allen, B. D. (2006). *Concept map scoring: Empirical support for a truncated joint poisson and conway-maxwell-poisson distribution method.*

Ausubel, D. P. (1968). *Educational psychology: a cognitive view.* Holt, Rinehart and Winston: New York.

Baayen, R. H., Piepenbrock, R., & Gullikers, L. (1995). *The CELEX lexical database (release 2).* [CD]. Philadelphia, PA: Linguistic Data Consortium, University of Pennsylvania.

Baddeley, A. D. (1986). *Working memory*. Oxford: Clarendon Press.

Binder, J. R., Westbury, C. F., McKiernan, K. A., Possing, E. T., & Medler, D. A. (2005). Distinct brain systems for processing concrete and abstract concepts. *Journal of cognitive neuroscience*, *17*(6), 905-917.

Burmistrov, I. V., & Shmeliov, A. G. (1992). Exsort and DCS: Prototype program shells using a psychosemantic approach to concept acquisition, representation and assessment. In P. Brusilovsky & V. Stefanuk (Eds.), *Proceedings of the east-west conference on emerging computer technologies in education.* (p. 46-51).

Carlson, N. R., Buskist, W., Enzle, M. E., & Heth, C. D. (2005). *Psychology: the science of behaviour* (3rd Canadian ed.). Toronto: Pearson Allyn & Bacon.

Coltheart, M. (1981). The MRC psycholinguistic database. *The Quarterly Journal of Experimental Psychology Section A*, *33*(4), 497-505.

Crutch, S. J., & Warrington, E. K. (2005). Abstract and concrete concepts have structurally different representational frameworks. *Brain: A Journal of Neurology*, *128*(3), 615-627.

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior*, *19*(4), 450-466.

Dixon, P., LeFevre, J., & Twilley, L. C. (1988). Word knowledge and working memory as predictors of reading skill. *Journal of educational psychology*, *80*(4), 465-472.

Forster, K. I., & Hector, J. (2002). Cascaded versus noncascaded models of lexical and semantic processing the turple effect. *Memory and Cognition*, *30*(7), 1106-1116.

Freeman, L. A., & Jessup, L. M. (2004). The power and benefits of concept mapping: measuring use, usefulness, ease of use, and satisfaction. *International Journal of Science Education*, *26*(2), 151-169.

Gabora, L., Rosch, E., & Aerts, D. (2009). Toward an ecological theory of concepts. *Ecological Psychology*, *20*(1), 84-116.

Gray, P. (2002). *Psychology*. New York: Worth Publishers.

Gunter, T. C., Wagner, S., & Friederici, A. D. (2003). Working memory and lexical ambiguity resolution as revealed by erps: A difficult case for activation theories. *Journal of cognitive neuroscience*, *15*(5), 643-657.

Hampton, J. A., & Moss, H. E. (2003). Concepts and meaning: Introduction to the special issue on conceptual representation. *Language and Cognitive Processes*, *18*(5-6), 505-512.

Hannon, B., & Daneman, M. (2001). A new tool for measuring and understanding individual differences in the component processes of reading comprehension. *Journal of Educational Psychology*, *93*(1), 103-128.

Johnson-Laird, P. N. (1987). The mental representation of the meaning of words. *Cognition*, *25*(1-2), 189-211.

Johnson-Laird, P. N., Herrmann, D. J., & Chaffin, R. (1984). Only connections: A critique of semantic networks. *Psychological bulletin*, *96*(2), 292-315.

Komatsu, L. K. (1992). Recent views of conceptual structure. *Psychological bulletin*, *112*(3), 500-526.

Medin, D. L., & Lynch, E. B. (2000). Are there kinds of concepts? *Annual Review of Psychology*, *51*(1), 121.

Murphy, G. L., & Medin, D. L. (1985). The role of theories in conceptual coherence. *Psychological review*, *92*(3), 289-316.

Nguyen, S. P. (2007). Cross-classification and category representation in children's concepts. *Developmental psychology*, *43*(3), 719-731.

Novak, J. D., Gowin, D. B., & Johansen, G. T. (1983). The use of concept mapping and knowledge via mapping with junior high school science students. *Science Education*, *67*(5), 625-645.

Oberauer, K., & Kliegl, R. (2006). A formal model of capacity limits in working memory. *Journal of Memory and Language*, *55*(4), 601-626.

Paivio, A. (1991). Dual coding theory: Retrospect and current status. *Canadian Journal of Psychology/Revue Canadienne de Psychologie*, *45*(3), 255-287.

Rosch, E., & Mervis, C. B. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive psychology*, *7*(4), 573-605.

Rosenberg, S., & Kim, M. (1975). The method of sorting as a data-gathering procedure in multivariate research. *Multivariate Behavioral Research*, *10*(4), 489-502.

Ruiz-Vargas, J., Cuevas, I., & Marschark, M. (1996). The effects of concreteness on memory: Dual codes or dual processing? *European Journal of Cognitive Psychology*, *8*(1), 45-72.

Safayeni, F., Derbentseva, N., & Cañas, A. J. (2005). A theoretical note on concepts and the need for cyclic concept maps. *Journal of Research in Science Teaching*, *42(7)*, 741 - 766.

Samson, D., & Pillon, A. (2004). Orthographic neighborhood and concreteness effects in the lexical decision task. *Brain and language*, *91*(2), 252-264.

Schmiedek, F., Hildebrandt, A., Lovden, M., Wilhelm, O., & Lindenberger, U. (2009). Complex span versus updating tasks of working memory: The gap is not that deep. *Journal of Experimental Psychology: Learning, Memory, and Cognition.*, *35*(4), 1089-1096.

Schwanenflugel, P. J., & Stowe, R. W. (1989). Context availability and the processing of abstract and concrete words in sentences. *Reading Research Quarterly*, *24*(1), 114-126.

Stephenson, W. (1953). *The study of behavior: Q-technique and its methodology*. Chicago: University of Chicago Press.

Trochim, W. M. K. (1989). An introduction to concept mapping for planning and evaluation. *Evaluation and Program Planning*, *12*(1), 1-16.

Trochim, W. M. K., & Linton, R. (1986). Conceptualization for evaluation and planning. *Evaluation and Program Planning*, *9*.

Tufte, E. R. (1991). *Envisioning information* (Rev. ed.). Cheshire, Conn.: Graphics Press.

Wagner, S., & Gunter, T. C. (2004). Determining inhibition: Individual differences in the ”lexicon context” trade-off during lexical ambiguity resolution in working memory. *Experimental Psychology*, *51*(4), 290-299.

West, D. C., Pomeroy, J. R., Park, J. K., Gerstenberger, E. A., & Sandoval, J. (2000). Critical thinking in graduate medical education. *Journal of the American Medical Association*, *284*(9), 1105.

Appendix A

Table 14
Items used in Experiment 1

Concept
Intermittent reinforcement
Secondary reinforcement
Appetitive stimulus
Aversive stimulus
Chaining
Consequence
Discriminative stimulus
Extinction
Negative reinforcer
Operant response
Place Learning
Positive Reinforcer
Punishment
Reinforcement schedules
Response rate
Spontaneous recovery

Note. The first two concepts were defined in Carlson et al., 2005, and the remaining were defined in Gray, 2002.

Appendix B

Table 15
Word lists used for Concrete nouns within each block of Experiment 2

3 Words		5 Words		7 Words		9 Words	
Random	Structured	Random	Structured	Random	Structured	Random	Structured
LEAF	LEAF	LEAF	LEAF	LEAF	LEAF	LEAF	LEAF
YACHT	YACHT	YACHT	YACHT	YACHT	YACHT	YACHT	YACHT
RECORD	RECORD	RECORD	RECORD	RECORD	RECORD	RECORD	RECORD
		POLE	POLE	POLE	POLE	POLE	POLE
		PLANET	PLANET	PLANET	PLANET	PLANET	PLANET
				PLIERS	PLIERS	PLIERS	PLIERS
				LEMON	ORANGE	LEMON	ORANGE
						SUNSET	SUNSET
						NAPKIN	AUTUMN

Table 16
Word lists used for Abstract nouns within each block of Experiment 2

3 Words		5 Words		7 Words		9 Words	
Random	Structured	Random	Structured	Random	Structured	Random	Structured
SOUL	SOUL	SOUL	SOUL	SOUL	SOUL	SOUL	SOUL
GUESS	GUESS	GUESS	GUESS	GUESS	GUESS	GUESS	GUESS
BELIEF	BELIEF	BELIEF	BELIEF	BELIEF	BELIEF	BELIEF	BELIEF
		TALE	TALE	TALE	TALE	TALE	TALE
		TENURE	TENURE	TENURE	TENURE	TENURE	TENURE
				MALICE	MALICE	MALICE	MALICE
				TOPIC	VIRTUE	TOPIC	VIRTUE
						FACTOR	FACTOR
						HYBRID	THEORY

Appendix C

Paragraphs

Experimental paragraphs that constitute the reading comprehension test administered in both experiments.

Practice. A GATH resembles a SHET but is heavier. A SHET resembles a COUCH but is heavier. A MUNT resembles a LAMP but is heavier.
heavier - gath > shet > couch > lamp; munt > lamp

Block 1. A SAMP resembles a BERL but is slower and weighs more. A NORT resembles a JET but is faster and weighs more. A BERL resembles a CAR but is slower and weighs more.
speed - nort > jet > car > berl > samp
weight - nort > jet > car; samp> berl > car

Block 2. A DERP resembles a PINE but is taller and lives longer. A BUFT resembles a PETUNIA but is shorter and lives longer. A ROSP resembles the DERP but is taller and lives longer.
height - rosp > derp > pine > petunia > buft
lifespan - rosp > derp > pine > petunia; buft > petunia

Block 3. A FILP resembles a COFT but is smaller, has a longer neck, and nests on land. A COFT resembles a ROBIN but is smaller and has a longer neck. A MIRT resembles an OSTRICH but is larger and has a longer neck.
size - mirt > ostrich > ROBIN > coft > filp
neck length - mirt > ostrich > robin; filp > coft > robin
nests on land - filp, ostrich
doesn’t nest on land - coft, robin

Block 4. A MARB resembles a BUTTERFLY but is more colorful and larger. A JERP resembles an ANT but is less colorful and larger. A TOLP resembles a MARB but is more colorful, larger and lives in a colony.
colorful - tolp > marb > butterfly > ant > jerp
size - tolp > marb > butterfly > ant; jerp > ant
lives in colonies - tolp, ant
doesn’t live in colonies - marb, butterfly

Block 5. A LORK resembles a TILN but is shorter, eats more, and lives on land. A long-legged TILN resembles a MONKEY but is shorter and eats more. A WEMP resembles a GIRAFFE but is taller and eats more.
height - wemp > giraffe > monkey > tiln > lork
amount eaten - wemp > giraffe > monkey; lork > tiln > monkey
long-legged - tiln, giraffe
short-legged - monkey
lives on land - lork, giraffe
doesn't live on land - tiln, monkey

Block 6. A GORT resembles a TESK but is larger, sweeter, and grows on the vine. A PESH resembles a LEMON but is smaller and sweeter. A TESK resembles a WATERMELON but is larger and sweeter.
size - gort > tesk > watermelon > lemon > pesh
sweetness - gort > tesk > watermelon > lemon; pesh > lemon
round - gort, tesk
not round - watermelon, lemon
grows on vine - watermelon
doesn't grow on vine - tesk, lemon

True/False Statements

All of the True/False statements are available in Hannon and Daneman (2001). However, one statement representing each sub component will be presented.
Text-based memory - A GORT is larger than a TESK.
Text-based inferencing - A TILN doesn't live on land.
Knowledge integration - Like HONEYBEES, MARBS fly in the air.
Knowledge access - A BLUEJAY lives in canada, whereas an OSTRICH typically doesn't.