

ANALYSIS OF THE HIERARCHICAL NATURE OF CLINICIANS'
ORGANIZATION OF MENTAL DISORDERS

Except where reference is made to the work of others, the work described in this thesis is my own or was done in collaboration with my advisory committee. This thesis does not include proprietary or classified information.

Jared Wayne Keeley

Certificate of Approval:

Jeffrey Katz
Assistant Professor
Psychology

Roger K. Blashfield, Chair
Professor
Psychology

Alejandro A. Lazarte
Assistant Professor
Psychology

Stephen L. McFarland
Dean
Graduate School

ANALYSIS OF THE HIERARCHICAL NATURE OF CLINICIANS'
ORGANIZATION OF MENTAL DISORDERS

Jared Wayne Keeley

A Thesis
Submitted to
the Graduate Faculty of
Auburn University
in Partial Fulfillment of the
Requirements for the
Degree of
Master of Science

Auburn, Alabama
August 8, 2005

ANALYSIS OF THE HIERARCHICAL NATURE OF CLINICIANS'
ORGANIZATION OF MENTAL DISORDERS

Jared Wayne Keeley

Permission is granted to Auburn University to make copies of this thesis at its discretion, upon request of individuals or institutions and at their expense. The author reserves all publication rights.

Signature of Author

Date

Copy sent to:

Name

Date

THESIS ABSTRACT
ANALYSIS OF THE HIERARCHICAL NATURE OF CLINICIANS'
ORGANIZATION OF MENTAL DISORDERS

Jared Wayne Keeley
Masters of Science, August 8, 2005
(B.S. Knox College, 2003)

65 Typed Pages

Directed by Roger K. Blashfield

The organization of mental disorders by clinicians can be viewed as a type of folk taxonomy. If this is true, the organizations of clinicians should exhibit certain properties; specifically, they should be hierarchical. This hierarchical nature is implicit in the Diagnostic and Statistical Manual of Mental Disorders – Fourth Edition (DSM-IV), which is the current organizational standard in the field of mental health. This study examined the organizations of three samples of clinicians: two expert samples (N = 7 and 21, respectively) and one novice sample (N = 13). The results indicate that clinicians do organize mental disorders in a hierarchical fashion, but not to the degree found in the DSM-IV. Remarkably, clinicians of varying experience and geographic location tested under separate methodological conditions produced the same hierarchical pattern.

ACKNOWLEDGMENTS

The author would like to thank Roger Blashfield for his unwavering support and assistance in the development and implementation of this thesis project. Thanks to Elizabeth Flanagan for the use of her previously collected data and for her insightful and helpful comments on previous drafts of the work. Jeffery Katz and Alejandro Lazarte also deserve a hearty thanks for their assistance with theoretical and statistical issues involved in the development of this thesis, as well as their comments in the formative stages of the project. A special thanks goes to Mei Jang for her assistance in implementing and constructing the computer programs used in the study. Thanks also to Danny Burgess, David Everett, and Andy Cohen for their comments and support throughout the process.

Style manual used APA Publication Manual (5th edition)

Computer software used Microsoft Word 2002

TABLE OF CONTENTS

LIST OF TABLES	viii
LIST OF FIGURES	ix
I. INTRODUCTION	1
II. METHOD.....	9
III. RESULTS.....	19
IV. DISCUSSION.....	32
V. REFERENCES.....	44
VI. APPENDICES.....	50
A. Mental disorder stimuli	
B. Hierarchical organization of the DSM-IV	
C. Tables	
D. Figures	

LIST OF TABLES

1. Means and standard deviations across levels of the hierarchy for the combined sample by methodology.....	55
2. Means and standard deviations across levels of the hierarchy for the expert samples by methodology.....	56
3. Means and standard deviations across levels of the hierarchy for the novice (Auburn) sample by methodology.....	57
4. Correlations of the sortings of the Georgia sample participants with the DSM	58
5. Correlations of the sortings of the Alabama sample participants with the DSM	59
6. Correlations of the sortings of the Auburn sample participants with the DSM	60

LIST OF FIGURES

1. Mean distances within groups across levels of the hierarchy for the combined samples by methodology.....	62
2. Mean distances within groups across levels of the hierarchy for the expert samples by methodology.....	63
3. Mean distances within groups across levels of the hierarchy for the novice sample by methodology.....	64
4. Comparison of the mean distances within groups across expert status and methodology.....	65

Analysis of the Hierarchical Nature of Clinicians' Organization of Mental Disorders

The area of study known as folk taxonomy has grown in popularity both in cognitive (Atran, 1995; Bailenson, Shum, Atran, Medin, & Coley, 2002; Lopez, Atran, Coley, Medin, & Smith, 1997; Medin, Lynch, Coley, & Atran, 1997) and clinical psychology (Flanagan, 2003; Flanagan & Blashfield, 2000). A folk taxonomy is a hierarchical organization of concepts developed within a culture. Typically, each culture has its own organizing principles and ways of classifying a group of similar categories, such as living things. One would expect that these organizations would be as varied as the cultures that produce them, as different cultures exist in different environments, speak different languages, have different lifestyles, etc. However, anthropological work (e.g. Berlin, 1992) has shown that there are remarkable similarities in the ways different cultures organize living things, and that these organizations retain some features of “scientific” classifications, although they often exhibit a simpler structure.

If it is true that people classify objects in similar ways across cultures, then perhaps there is some basic cognitive process which underlies this classification that is common to all humans. Many cognitive psychologists have eagerly examined this possibility (Atran, 1995; Bailenson et al., 2002; Johnson & Mervis, 1997; Lopez et al., 1997; Medin et al., 1997; Rosch, 1978). These psychologists

have developed methods for examining and testing the folk taxonomies of different cultures.

For example, Atran and colleagues (Atran, 1995; Bailenson et al., 2002; Lopez et al., 1997) have examined how the Itza Maya Indians organize their classification of living things. The most striking characteristic of these folk taxonomies is their similarity, despite the fact that they have been studied across varying cultures. Specifically, the study by Bailenson et al. (2002) found that the classifications of U.S. taxonomic experts were closer to those of the Itza than to U.S. novices.

Medin et al. (1997) examined the organization of trees among different groups of tree experts in an effort to understand some of the influences that shape the taxonomies. They did this by having taxonomists, landscape workers, and park maintenance personnel sort cards with the names of different trees into similar groups based on their experience in working with trees (and hence a “folk” classification). Once the participants had done this, they were asked to combine the groups into larger groups until they were no longer able to do so. The experimenters then restored the original groups, and participants were asked to break those into smaller groups, again until they were no longer able to do so. The result produced a hierarchy which represented the organization and classification of trees for each participant. The current scientific organization of trees is hierarchical, which is meant to mimic the speciation process of evolution. By quantifying the participants’ hierarchies, the authors were able to compare the

“folk taxonomies” of the various tree experts to the accepted scientific classification.

Medin et al. found that these folk taxonomies did correspond to the scientific classification in many ways. However, the group with the most expertise in classifying trees (the taxonomists) was able to differentiate groups at every level of the hierarchy. In other words, the taxonomists meaningfully separated groups at every level and provided a good match to the scientific hierarchy. However, the landscape and maintenance workers did not reproduce the scientifically defined groups at higher, more abstract levels of the hierarchy, but fared just as well as the experts at lower, more specific levels. In other words, the landscape and maintenance personnel were able to distinguish a White Oak from a Northern Red Oak from a European Black Alder, but were unable to distinguish more abstract groupings, often just lumping categories together like “conifers, birches, and the rest.” This study demonstrates that, in general, the ways in which these groups functionally organize trees on a daily basis corresponds fairly well to the actual scientific taxonomy, but those with more scientific training produce more differentiated organizations.

The various versions of the *Diagnostic and Statistical Manual of Mental Disorders* (DSM) have been conceptualized as a form of folk taxonomy specific to the “culture” of psychiatry and psychology (Flanagan, 2003; Flanagan & Blashfield, 2000). In other words, authors such as Flanagan and Blashfield hypothesize that the organization of the concepts of mental disorders as presented in the DSM is practically useful to clinicians in the same way as an

organization of living species is useful for tree experts. Therefore, the methods that have been used to study folk taxonomical structures can be used to study clinicians' organizations of mental disorders.

Flanagan (2003) adapted the methods used by Medin et al. (1997) to study the taxonomic structure of mental disorders as seen by clinicians. Her rationale was to examine the ways in which clinicians categorize mental disorders and to determine if any similarities arose in those categorizations. The general procedure required participants to sort a series of mental disorders into meaningful categories. However, rather than having participants try to reproduce the "scientific" classification (i.e. the DSM), Flanagan instructed the clinicians to sort the disorders based on their experience with treating the disorders, in order to tap their "folk" categorization. She tested subjects across two varieties of this methodology: a progressive grouping sorting and a tabletop sorting. In the progressive grouping method, clinicians first sorted the disorders into groups, and then were told to combine the groups, and combine them again, until they were no longer able to do so. The original groups were then recovered, and the clinicians were told to break the groups apart, again until they were no longer able to do so. This created a hierarchy of mental disorders, and mirrored the methodology used by Medin et al. (1997). In the tabletop sorting, clinicians took the same set of mental disorders, and again initially sorted them into groups. However, this time they sorted the groups into a two-dimensional space (a table top), so that the distance between groups was meaningful (i.e. similar groups were close together and different groups were far apart).

Among her findings, Flanagan found that there was a high level of cultural consensus among the clinicians (see Romney, Weller, & Batchelder, 1986 for a method of measuring cultural consensus). However, working backwards, there was no clear taxonomic organization that arose from the consensus of the sortings. Further, anecdotally, she reported that clinicians found the dimensional sorting task easier and more meaningful, suggesting that perhaps clinicians' conceptualization of mental disorders fit better with a dimensional model.

The current study

Flanagan's work left several key questions unanswered. Part of the original aim of the work was to achieve an understanding of the way in which clinicians organize mental disorders. In other words, when left to rely on their own clinical experience (thereby not relying on the established categorization of the DSM), how do clinicians organize and classify the various types of psychopathology they encounter? The DSM organizes mental disorders into a hierarchy, but do clinicians think about disorders in a hierarchical manner? Others have argued that mental disorders may be better represented along a series of dimensions (e.g. Costa & Widiger, 1994). Aside from the question of how mental disorders actually exist in the state of nature (if such a question can even be answered), it is useful to know how clinicians think about the array of psychopathology before them in the hopes of making the current classification system more utilitarian and cognitively efficient. Hence, Flanagan (2003) adopted the methodology of Medin et al. (1997) in an attempt to examine the structure of clinicians' folk taxonomies.

However, Flanagan (2003) did not analyze the data in such a way as to examine the structure of the classifications produced. Rather, she was interested in the clusters of disorders that were produced, and if these corresponded to the classification of the DSM. She did note that several of the dimensional sortings in her data set looked particularly hierarchical, but she had no means of empirically evaluating that statement. The current work is an attempt to further our understanding of that structure through re-analyzing the same data sets collected by Flanagan. In the current study, the data will be analyzed so as to examine to what degree participants' sortings are hierarchical or not.

Further, Flanagan used two methodologies for sorting: a progressive grouping method and a tabletop method, which supposedly reflect organizing mental disorders in a hierarchical and dimensional fashion, respectively. It may be that each methodology pulls for its own particular kind of classification. However, the progressive grouping methodology might or might not produce classifications that are more hierarchical and vice versa. Clinicians' own folk organizations may overwhelm the effects of the methodology, and appear hierarchical in both methods, or in neither. The current study will compare sortings across the types of sorting method used to examine the potential effects of the methodology on the shape of the resulting data.

Also, Flanagan included subjects of varying levels of expertise, as a person's level and type of experience could have an effect on the person's classifications (Medin et al., 1997). Specifically, Medin et al. found that all

participants were able to differentiate groups meaningfully at the lower, more specific levels of the hierarchy. However, at higher, more abstract levels, only experts were able to make distinctions between groups. Therefore, the current study will examine if the participant's level of expertise has an effect on the specificity of the sorting achieved. However, Medin et al. used "real" stimuli for their sorting that had a predetermined correct scientific classification. The stimuli used by Flanagan, on the other hand, are ill-defined scientifically. The nature of the stimuli may have created the effect of expertise found in the Medin et al. study. For example, groups such as the Northern Red Oak and European Black Alder are defined by perceptually distinct characteristics, and therefore easy to separate at a low level of a hierarchy. However, the distinctions between higher level groups in the scientific organization of trees are not characteristics that are readily perceived (such as leaf shape, bark texture, etc.). Distinctions at these upper levels are based upon evolutionary history, which may not be readily evident to a landscaper or maintenance worker. However, mental disorders are defined in terms of symptoms that often lack specific, perceptual distinctions (i.e. a schizophrenic patient does not necessarily "look" different than a borderline patient, and their symptoms may overlap or appear very similar in expression). Nonetheless, the conceptual distinctions between these groups are very apparent to clinicians (i.e. schizophrenics are etiologically and conceptually very different from borderlines).

Hypotheses

- 1) Clinicians' sortings will not demonstrate a hierarchical structuring, regardless of the type of methodology used. While Berlin (1992) has argued that hierarchical structuring is a natural cognitive process, and so should be evident in most any natural classification system, I argue that the current state of knowledge regarding mental disorders is too ill-defined for any clear structure to emerge.
- 2) Classifications made using the progressive grouping methodology will appear more hierarchical than those produced in the tabletop methodology. I predict that the nature of the progressive grouping task is forcing clinicians to think in a hierarchical fashion, and so the progressive sorting results will appear more hierarchical than those produced in the tabletop methodology.
- 3) All clinicians will be able to differentiate groups at higher, abstract levels but only those with high levels of expertise will be able to make meaningful distinctions at lower, more specific groupings of disorders. This is counter to Medin et al.'s finding of experts differentiating more levels than novices. However, the stimuli used in the Medin et al. study were defined concretely, and are therefore perceptually distinct. Mental disorders are abstract and loosely defined, often with much overlap between the categories (Hall, 1996; Kessler, 1995). Therefore, I predict that specific distinctions will be difficult for participants, and so only those with a high level of experience with mental disorders will be able to differentiate groups at a specific level.

4) All classifications will have little resemblance to the structure of the DSM. While the DSM is the “scientific standard” in use in the field, I predict that clinicians will produce sortings only loosely related to the DSM. Many clinicians, researchers, and theorists have proposed alternative models to the DSM, and many clinicians ascribe to such models. For example, there are dimensional models of personality disorders and related Axis I disorders (Costa & Widiger, 1994; Lynam & Widiger, 2001; Miller, Reynolds, & Pilkonis, 2004), the tripartite model of anxiety and depression (Brown, Chorpita, & Barlow, 1998; Clark & Watson, 1991; Zinbarg & Barlow, 1996), and models of substance use disorders (Langenbucher et al., 2000) that all propose different structures than those seen in the DSM.

Method

Participants

There were three groups of subjects included in this study: (1) Georgia expert clinicians, (2) Alabama expert clinicians, and (3) Auburn University graduate student clinicians. Data for all three groups were previously gathered as part of Elizabeth Flanagan’s dissertation (2003).

The Georgia Clinicians This group of participants was solicited at the April 2001 meeting of the Georgia Psychological Association. The experimenter sat at a centrally located table which displayed a sign reading “Please participate in a study of clinicians.” As compensation, participants were asked if they would like to be entered in a drawing for a \$30 gift certificate redeemable for APA books.

Seven clinicians agreed to participate in the study. Flanagan (2003) initially defined this as an “expert” sample, and so clinicians were only included if they had at least 10 years of experience. However, due to the small number of participants in this group, all seven participants were included in the analyses. The participants had an average of 15 years ($SD=8.0$) of experience since seeing their first client and an average age of 43 ($SD=9.9$). One participant was Hispanic, while the rest were Caucasian. Two participants were male and five were female.

The Alabama clinicians This group of clinicians was in attendance at the annual Alabama Psychological Association meeting in June of 2001. They were solicited for participation by the experimenter who sat at a centrally located table with a sign reading “Please participate in a study of clinicians.”

Thirty-one clinicians agreed to participate in the study. Of these, 21 participants’ data were suitable for analysis. As this was an “expert” sample, clinicians were only included if they had at least 10 years of experience. Seven participants did not meet this criterion, and so were dropped. Additionally, one participant was dropped because he had a master’s degree where all other participants had doctorates; another was deleted because she was a doctor of internal medicine and not a psychologist; and a final person was dropped because she included only 9 of 67 disorders in the sorting task (see below for an explanation of the task). The remaining participants had a mean experience of 24.3 years ($SD=10.5$). The average age of the participants was 50.8 ($SD=10.2$).

Fourteen participants were female, and one participant was non-Caucasian (Flanagan, 2003).

Auburn Graduate Clinicians This group of participants was solicited from two separate graduate departments at Auburn University: the clinical psychology and the counseling psychology graduate programs. Participants were approached if they had at least one year of clinical experience. Clinical psychology students were approached directly by the experimenter where counseling psychology students were notified of the study through a posting in their mailroom and an e-mail sent to every student containing an electronic copy of the same posting. Students were compensated \$25 if they participated two times and \$10 if they participated once.

Of the 24 clinical psychology students who were eligible to participate in the study, all 24 agreed to participate. Of the 20 eligible students in the counseling program, however, only 5 agreed to participate, and four of the five counseling graduate students were personal acquaintances of the experimenter. The average amount of experience for the combined sample was 3.9 years ($SD=1.8$) and the average age was 28.8 ($SD=3.8$; Flanagan, 2003). Unfortunately, at the time of this study, photographic copies of the sortings from this sample had been lost, and so only 14 participants' data were suitable for analysis. One of these participants included insufficient data in the sorting for calculating all levels of the hierarchy (see below for an explanation of this calculation), and so was dropped for a total of 13 participants in the Auburn sample. Of these 13, the average years of experience was 4.1 ($SD=1.6$) and the

average age was 28.1 ($SD=1.6$). Of the 13, four were male and three were counseling students.

Stimuli

The stimuli used in all three samples were names of mental disorders printed on 3x5 inch index cards. The names of disorders were selected to be representative of diagnoses commonly used by mental health professionals.

There are over 400 diagnostic categories included in the DSM-IV. Participants would be overwhelmed both in term of time commitment and cognitive effort to sort 400 categories, thus Flanagan selected only a subset of these disorders. Flanagan (2003) outlined the selection process, which is reproduced here. Two of the authors of the DSM-IV (APA, 1994; as well as DSM-IV-TR, APA, 2000), Frances and First, published a book in 1998 entitled *Am I Okay?*, that was designed to be a lay-person's guide to mental disorders. Frances was the chairperson of the DSM-IV task force, and so was responsible for the overall organization of the text. First was the editor of the DSM-IV text, and was responsible for the final wording and formulation of the narrative and diagnostic criteria. Each of their positions afforded them considerable control over the nature and structure of the DSM-IV. Blashfield (2001) stated that their popular book, *Am I Okay?* seems to represent Frances and First's summary of the most important disorders in the DSM-IV.

Frances and First (1998) included 67 disorders in their book. These disorders included a representative of each of the major higher order categories present in the DSM-IV, and encompassed a similar breadth of psychopathology.

Therefore, these 67 disorders are a representative sample of mental disorders chosen by authors of some repute, and as 67 is a much more manageable number for a sorting task than the 400 diagnoses in the DSM-IV, Flanagan selected these disorders as the stimuli for the sorting task.

The name of each disorder was written on a 3x5 index card. These cards served as the stimuli for the sorting task. A list of the disorders is included in Appendix A as they appeared on the card. The names of the disorders were for the most part identical to those used by Frances and First. However, some names were altered to correspond to common language use of the term. For example, instead of displaying Posttraumatic Stress Disorder on the card, participants saw "PTSD," which is a commonly used abbreviation. Similarly, the personality disorders were not followed by the words "personality disorder" (e.g. antisocial personality disorder), but were presented as only a short description (e.g. antisocial), as to avoid clinicians sorting these disorders together because the last two words were identical.

Procedure

Two different procedures were used: a progressive grouping task and a tabletop sorting task. The Georgia sample completed the progressive grouping methodology, where the Alabama sample completed the tabletop sorting. The members of the Auburn sample completed the procedure twice, in counterbalanced conditions. However, as some of the data for this sample had been lost, it was not possible to include comparisons between the two testing times as only a few participants had data remaining at both times. Therefore,

only data from the first administration will be used, as this administration can be considered independent and equivalent to the conditions of the other samples. Eight participants completed the progressive grouping methodology and five completed the tabletop sorting in the first sorting task.

The Alabama and Auburn samples were tested on the full set of 67 cards. However, the Georgia sample was tested on a set of 66, as one disorder (substance induced disorder) was not included. For the sake of equivalence, substance induced disorder was dropped from the sortings of those clinicians which included the disorder in the Alabama and Auburn samples. Therefore, all further analyses will be based on only the 66 disorders included in all samples. Participants were allowed to examine the entire deck and to discard any categories with which they were not familiar. If participants were uncertain whether to keep a particular category, they were told that they would be asked to sort the cards into groups that “felt the same” or that “had similar treatments,” and to include the card if they felt that they had sufficient knowledge or experience with the category to complete the task. The two methodologies are explained below.

Progressive grouping methodology

Participants sat at a 3x5 foot table directly across from the experimenter. They were allowed as much time as they wished to complete the task. Once clinicians had examined the stimuli and discarded unfamiliar diagnoses, they were read the following instructions:

Put together the diagnoses that have similar treatments into as many groups as you'd like. I am trying to determine the disorders that "feel" the same. I am not interested in what the DSM says. I am interested in what you've found from your clinical experience.

Once the participant had completed sorting the categories, the experimenter asked the participant to name each group, and these names were then recorded. The experimenter then asked the participant to combine the existing groups into larger groups based on similarity of treatment. This was repeated until the participant refused to combine any further groups, or until all disorders were placed in one group. Then, the participant's original sorting was restored, and the experimenter asked the participant to break these into smaller groups. Again, the procedure was repeated until the participant indicated that no further splits seemed appropriate. The experimenter recorded the participant's groupings as they were completed.

After the clinicians finished the task, they were given a demographic sheet which asked for their age, sex, race, years of clinical experience (starting at their first client), highest degree obtained, theoretical orientation, and how often they consulted the DSM. They were then thanked for their participation and dismissed.

Tabletop methodology

Once the clinicians had initially examined the deck of stimuli and discarded the disorders with which they were not familiar, they were read the following instructions:

Put together the diagnoses that have similar treatments into as many groups as you would like. I am trying to determine the disorders that “feel” the same. I am not interested in what the DSM says; I am interested in what you’ve found from your clinical experience. After you have created the groups, put the groups that have the most similar treatments next to each other. When you finish, the table should be a multi-dimensional space where the groups that have the most similar treatments are next to each other and the groups that have the most different treatments are farthest away from each other.

All clinicians were tested individually at a 3x5 foot table. The clinicians sat directly across from the experimenter. They were allowed as much room to work as the table allowed, and as much time as desired to complete the task.

The experimenter recorded the clinicians’ sortings *in vivo* in a drawing and took two photographs of the sorting after the participant had finished. As the physical distance of the groups was assumed to be meaningful, the dependent measure of distance/similarity was the physical distance between groups as measured in inches.

Further, after the participants completed their sorting, they were asked to name each of the groups, and the names were recorded on the experimenter’s drawing. Some participants in the Alabama sample spontaneously provided a rationale for their sorting as well (such as naming or describing the dimensional axes). Therefore, for the Auburn students who completed the tabletop sorting

(who were tested chronologically after the Alabama sample) were required to provide names to the dimensions as well as a rationale for the sorting.

After the clinicians finished the task, they were given a demographics sheet which asked for their age, sex, race, years of clinical experience (starting at their first client), highest degree obtained, theoretical orientation, and how often they consulted the DSM. They were then thanked and dismissed.

Analyses

Structuring of the data

The data initially were converted into a distance matrix. The matrix was a square grid of distance measures between all possible pairs of categories (having a distance of zero from itself). As the matrix was symmetric, or identical on each side of the diagonal, only the lower triangle was used to avoid artificially inflating any calculations. The distance created varied by the methodology used.

For the progressive grouping data, the distance is the number of nodes or levels of the hierarchy that separate two disorders. For example, if disorder A is grouped with disorder B at the lowest possible level used in the sorting, the two disorders will have a distance of one. However, if disorder A and disorder C are not together in that smallest group, but at the next level in the sorting their respective groups are joined, they will have a distance of two. The number of possible levels was standardized to seven, as this is the maximum number of levels included in the DSM-IV's implied hierarchy. In other words, if a participant only included five levels in her sorting, these five would be "stretched" into a possible seven levels by matching the specificity of the created levels as closely

as possible to those given in the DSM. For example, the level of individual disorders in the sorting was matched to the level of individual disorders in the DSM, and the highest level included in the sorting was linked to the highest level of differentiation in the DSM (i.e. the difference between Axis I and Axis II, see Appendix B). All other levels were then matched based upon the closest content correspondence of the groups in the sorting to the DSM. A computer program written in GW-BASIC generated the distance matrix from the grouped sorting of the participant.

For the tabletop methodology, the measure was the physical distance in inches between the cards in the sorting. As previously stated, the experimenter took a picture of each participant's final sorting. This picture was then divided into a grid, and the Cartesian coordinates for every disorder were entered into a second GW-BASIC program. This program then computed the Euclidean distance for every combination of disorders. As the maximum distance in each sorting was different, and thereby each participant's sorting had its own scale, all sortings were scaled to have a maximum distance of seven. Thence, the distance matrices of the progressive grouping data and tabletop data were standardized to the same scale with values ranging from zero to seven.

Not every participant used every card, as they were allowed to discard unfamiliar categories. As a result, there was a large amount of missing data in the matrix. One strategy to impute the missing data was to set all missing disorders to a distance one unit farther than found in the sorting (i.e. to a value of 8). The logic behind this choice was that a disorder not included by the

participant is further distant than any included. However, this strategy profoundly skewed all later analyses. For example, calculating correlations between distance matrices where large amounts of data all had the same value essentially pulled the correlation closer to zero, which was not representative of the relationship of the real data points. Therefore, all missing data were left as missing, and any combinations or comparisons of participants' data were weighted by the amount of data present in their sortings.

Results

Hypothesis 1: Overall hierarchical organization

The primary interest of this study is to examine the structure and organization of participants' sortings, specifically if the sortings follow a hierarchical organization or not. In a hierarchical organization, members of a group will be close to each other, but progressively further from other groups at each level of the hierarchy. For example, the hierarchical organization of the DSM-IV as taken from its table of contents is included in Appendix B for the 66 categories included in the analysis. In a grouping at the lowest possible level, necessarily all members of the group are only a distance of one unit from other members of the group, such as the grouping of ADHD, Conduct Disorder, and Oppositional/Defiant Disorder. However, at the next level of the hierarchy, there are more members of the group "childhood disorders," and we would expect the average distance of members of that group to approach (but not equal) 2. Therefore, the average distance among group members at each level of the hierarchy should become progressively greater at each level of the hierarchy in

the distance matrices of the participants if they are following a hierarchical organization. Groups, regardless of methodology used, were hypothesized to demonstrate no hierarchical organization, or, more concretely, this hypothesis can be translated to mean that the distances among groups at each level of the hierarchy were expected to be roughly equal.

This hypothesis was examined by calculating the average distance of groups of disorders for each of four levels of the hierarchical organization of the DSM-IV. The DSM is the accepted standard among mental health professionals in the United States. Therefore, the DSM represents the “agreed upon” organization that most closely represents the scientific organization of other fields, such as a hierarchical classification of living species in biology. The stimuli used in the sorting task represented only a portion of the hierarchy present in the DSM. Specifically, only four levels of the DSM were represented in the stimuli. The four levels that were present, moving from lowest to highest, correspond to individual disorders, subgroups of disorders (such as the clusters of personality disorders), major groups of disorders (mood disorders, substance disorders, etc.), and the major axes (Axis I versus Axis II disorders, corresponding to acute versus stable conditions). For example, the lowest level present in the DSM corresponds to subtypes and specifiers of disorders, such as Schizophrenia, Paranoid Type versus Schizophrenia, Catatonic Type. The stimulus representing this disorder was simply “Schizophrenia” and did not differentiate subtypes. Therefore, it was not possible to accurately represent all seven levels of the DSM hierarchy. Nonetheless, as noted before, the data were

standardized to a scale of seven. The data were left in this scale so that the full range of variability between groups of disorders would be represented.

For the four levels present given the constraints of the stimuli, I calculated an average distance among groups of disorders at each of the four levels for each participant. Again, these distances were weighted by how many data points were included in their calculation. For the purpose of this analysis, all participants were combined regardless of expert versus novice status to increase the sample sizes and power of the analyses (For an examination of differences between experts and novices, see Hypothesis 3 below). The mean distances and standard deviations for the two methodologies are shown in Table 1. Figure 1 shows the graphical relationship of these means by methodology.

In examining Figure 1, some levels of the hierarchy do have higher distances within their groups than others. In other words, if the degree of change between levels is significant, i.e. if the slope of the line between two levels does not equal zero, then there is hierarchical organization. Further, a perfect hierarchical organization would be equal to an increase of one average distance unit for each step in the hierarchy. However, this change does not necessarily correspond to a slope of 1, as the distances created in participants' sortings correspond to a seven step scale. Again, due to the constraints of the stimulus set, only four of the seven hierarchical levels of the DSM can be represented. The average distance within groups that would correspond to an exact hierarchical match to the DSM would be a distance of 2 for level 1, 3 for level 2, 5 for level 3, and 7 for level 4. Therefore, a slope of 1 between levels 1 and 2

corresponds to an exact hierarchy, where a slope of 2 between levels 2 and 3 and levels 3 and 4 would indicate the same amount of change as the slope of 1 between levels 1 and 2. The hierarchical structure of the data was tested in a series of specific contrasts in a MANOVA with each of the four levels entered as a dependent measure and the two methodologies entered as a between-subjects factor. I combined the participants in the three samples for the purpose of this analysis (see Hypothesis 3 below). The overall test resulted in a Wilks' $\lambda(4, 36) = 0.034, p < .001$, indicating that there are differences present between the variables. The methodologies used to create the sortings were tested separately.

First, I tested whether the degree of change between levels was equal to zero or not, or if there was any slope to the line segments between levels. For the progressive grouping methodology, there was a significant degree of change between levels 1 and 2 and levels 2 and 3; $F(1, 14) = 20.490, p < .001$ and $F(1, 14) = 54.004, p < .001$, respectively. However, there was no change between levels 3 and 4; $F(1, 14) = 1.359, p = .263$. The tabletop methodology followed the same pattern. There was a significant degree of change between levels 1 and 2, $F(1, 25) = 26.290, p < .001$, and between levels 2 and 3, $F(1, 25) = 125.519, p < .001$. There was no change between levels 3 and 4, $F(1, 25) = 3.555, p = .071$. Therefore, for both methodologies, there is some degree of hierarchical organization at the lower levels of the hierarchy, but no differentiation between the two highest levels (i.e. between Axes I and II).

While there is some degree of hierarchical organization, that organization may or may not be equal to the ideal. Again, a change of 1 unit of distance is expected between levels 1 and 2, where a change of 2 units of distance is expected between levels 2 and 3 and levels 3 and 4. The slope between levels 3 and 4 is equal to zero, so only the lower levels of the hierarchy were tested. For the progressive grouping method, the slope between levels 1 and 2 was equal to the ideal; $F(1, 14) = 0.131, p = .723$. However, the slope between levels 2 and 3 was not equal to the ideal; $F(1, 14) = 44.036, p < .001$. For the tabletop method, neither of the slopes between levels 1 and 2 and levels 2 and 3 were equal to the ideal slope; $F(1, 25) = 53.685$ and $F(1, 25) = 99.217$, both p 's $< .001$. Therefore, while the slopes at these levels are not equal to zero, only the slope between the low levels in the progressive grouping method was equal to the ideal slope. There is some level of hierarchical organization to the data, but that hierarchy is “weaker” than the hierarchy portrayed in the DSM.

Hypothesis 2: Differences due to methodology

I hypothesized that the nature of the methodology used would have an impact upon the structure of the sortings produced by clinicians. Specifically, I hypothesized that the progressive grouping methodology, which in structure resembles a hierarchical sorting, would produce sortings that appear more hierarchical than those produced under the tabletop sorting, which structurally resembles a dimensional viewpoint. Already, we have seen that both methodologies have some degree of hierarchical organization. However, it may be that the methods have differing degrees of that organization.

In examining Figure 1, two things seem apparent. First, the progressive grouping methodology is uniformly higher (i.e. more distant within groups) than the tabletop methodology. Second, the two lines for the progressive grouping and tabletop methodologies are largely parallel.

First, the progressive grouping methodology seems to have produced within group distances that are uniformly higher than those produced in the tabletop methodology. While this difference does appear large, it is meaningless. The absolute magnitude of distances within groups at each level are an artifact of the mathematical procedures used in their creation. For example, in the tabletop sortings, disorders could be placed in a stack at a given point in the sorting. This created groups of disorders with a distance of zero from each other (as they exist at the same Cartesian coordinate). Including these zero values in averages decreases the absolute possible magnitude of the average. However, in the progressive grouping methodology, there is no possible way for two disorders to have a distance of zero. Therefore, one cannot equate the magnitude of the distances created between the two methodologies. Further, the absolute magnitude of the distance within a group has no impact in and of itself on the determinacy of hierarchical organization. Only a level's magnitude relative to other levels has meaning in relation to hierarchical organization, as this concept is determined through slope. Therefore, the difference in height between methodologies observed in Figure 1 is uninterpretable and unimportant.

Second, the two lines appear to be parallel. This was tested by comparing the degree of change in distance between levels of the hierarchy across the two

methodologies (i.e. by comparing the slopes of the line segments between levels in Figure 1) in a specific contrast in the same MANOVA described under Hypothesis 1. The comparisons of level 2 with level 3 and level 3 with level 4 ($F(1, 39) = 0.002, p = .963$ and $F(1,39) = 0.856, p = .361$) show that we retain the hypothesis that the lines are parallel in these segments. However, the change between level 1 and level 2 is not equal between the two methodologies ($F(1,39) = 7.562, p = .009$), with the progressive grouping methodology evidencing a more substantial slope (refer to Figure 1). Between levels 1 and 2, both methodologies evidenced some degree of slope, but that this slope was near the ideal for the progressive grouping method (i.e. ≈ 1) while the slope for the tabletop method data was not ideal (see Hypothesis 1 above). Therefore, while both methods evidence some hierarchical organization between these levels, that organization is closer to the DSM in the progressive grouping method. Further, the difference between levels 2 and 3 is negligent (diff = .008 between slopes), indicating nearly perfect parallelism between methodologies at this level.

As these data for each methodology were aggregated across three samples, one may reasonably argue that differences between experts and novices within type of methodology may mask true differences between methodologies. Therefore, I also examined the differences between methodology types within only experts and only novices to remove potential biasing factors. The results of these analyses are presented next (For an examination of differences between experts and novices, see Hypothesis 3 below).

For the two expert samples, Georgia ($n = 7$) and Alabama ($n = 21$), the overall multivariate test indicated that there were differences within or between groups; Wilks' $\lambda(4,23) = .032, p < .001$. The mean distances and standard deviations within groups for each level of the hierarchy are presented in Table 2. These means are shown graphically in Figure 2. These two groups followed much the same pattern of results as the combined analyses. The differences between methodologies are uniform, creating parallel slopes between methodologies at each level; $F(1,26) = 1.166, p = .290$; $F(1,26) = 1.364, p = .253$; and $F(1,26) = .087, p = .770$.

For the Georgia sample, the slopes between levels 1 and 2 and levels 2 and 3 were significant; $F(1,6) = 9.186, p = .023$ and $F(1,6) = 47.241, p < .001$. The slope between levels 3 and 4 was not significant; $F(1,6) = .465, p = .521$. This pattern matches that of the combined sample described above.

The Alabama sample followed a similar pattern of results to that of the Georgia sample. The slopes between levels 1 and 2 and between levels 3 and 4 were significantly different from zero; $F(1,20) = 17.709, F(1,20) = 85.431$, both p 's $< .001$. There was no appreciable slope between levels 3 and 4; $F(1,20) = 1.705, p = .206$.

For both samples, only one slope was equal to the ideal. The slope between levels 1 and 2 in the Georgia sample was not significantly different from the DSM; $F(1,6) = 4.174, p = .087$. However, due to the small size of the Georgia sample, if more participants were included in this group, this difference between the progressive grouping and tabletop methodologies may disappear.

The novice samples follow a nearly identical pattern to that of the experts. There were 5 participants included in the tabletop method, and 8 included in the progressive grouping method. The mean distances and standard deviations within groups are shown in Table 3, with Figure 3 demonstrating the relationship graphically. The differences between methods were uniform across changes in level, with all slopes being parallel between the methodologies; $F(1,11) = 2.612$, $p = .134$, $F(1,11) = 2.707$, $p = .128$, and $F(1,11) = 2.369$, $p = .152$.

For both samples, the slopes at the lower levels were significantly different from zero, where the slope between levels 3 and 4 was flat. For the progressive grouping method, the tests for the slopes between levels 1 and 2 and between levels 2 and 3 were $F(1,7) = 14.372$, $p = .007$ and $F(1,7) = 17.681$, $p = .004$, respectively. The slope between levels 3 and 4 was equivalent to zero; $F(1,7) = .889$, $p = .377$. For the tabletop methodology, the tests for the slopes between levels 1 and 2 and between levels 2 and 3 were $F(1,4) = 9.284$, $p = .038$ and $F(1,4) = 91.486$, $p = .001$. Again, the slope between levels 3 and 4 was non-significant; $F(1,4) = 1.913$, $p = .239$.

Of all the significant slopes, only the slope between levels 1 and 2 of the progressive grouping method was equal to the DSM; $F(1,7) = .446$, $p = .526$. All others were not equal to the DSM. However, again this may be an artifact of limited sample size in the tabletop method. If more participants were in this group, and consequently the test had more power, the results between the two groups may be equivalent.

Hypothesis 3: Differences between experts and novices

I hypothesized that all clinicians would be able to make meaningful distinctions between groups at higher levels of the hierarchy, but that only experts would be able to make distinctions at lower levels of the hierarchy. To test for differences between experts and novices, the data that were aggregated in the analyses above were separated by sample and type of methodology used. Thereby, differences between experts and novices tested under the progressive grouping method were compared using data from the Georgia ($n = 7$) and Auburn ($n = 8$) samples. Differences in expertise in the tabletop method were tested with the Alabama ($n = 21$) and Auburn ($n = 5$) samples. The Auburn sample was split by the type of methodology used for each participant.

Means and standard deviations for the four groups can be found in Tables 2 and 3, respectively. The combination of the four samples is displayed graphically in Figure 4. Examining Figure 4 reveals that experts and novices both follow the same overall pattern described in the combined data. Remarkably, there seems to be little difference between the experts and novices, except for a few potential points of departure. To quantitatively test these differences, I conducted a MANOVA with the four levels of the hierarchy entered as dependent measures and the four samples entered as a fixed independent factor. The overall test indicated that there are differences between the variables; Wilks' $\lambda (12, 90.247) = .021, p < .001$. For specific contrasts, the methodologies were tested separately.

For the progressive grouping methodology, there were no differences in average distance between experts and novices at any level of the hierarchy; $F(1,13) = .063, p = .805$; $F(1,13) = 2.250, p = .158$; $F(1,13) = .994, p = .337$; and $F(1,13) = .678, p = .425$, respectively. Further, the slopes between all levels were equivalent; $F(1,13) = 2.501, p = .138$; $F(1,13) = 1.377, p = .262$; and $F(1,13) < .001, p = .986$.

For the tabletop methodology, again there was one difference in mean level distance between experts and novices. Levels 1, 2, and 3 were equivalent ($F(1,24) = .825, p = .373$; $F(1,24) = .172, p = .682$; and $F(1,24) = 1.995, p = .171$), where there was a difference in level 4 ($F(1,24) = 11.594, p = .002$). In more practical terms, the novices held the personality disorders and mental retardation as separate from Axis I groups, where the experts did not. However, all the slopes between levels were equivalent across the two groups; $F(1,24) = .390, p = .538$; $F(1,24) = 3.025, p = .095$; and $F(1,24) = 3.053, p = .093$. Again, these calculations are based upon two groups, one of which has a size of only five members. There may be a true difference between novices and experts in this area, with novices differentiating a hierarchical structure between Axis I and Axis II disorders (where every other sample did not). However, more likely, this difference is due to random fluctuation in a small sample, and would not be replicated.

Hypothesis 4: Resemblance to the DSM

I hypothesized that there would be little resemblance between participants' sortings and the structure seen in the DSM. While there does seem to be some

hierarchical organization in the lower levels of the data, this does not mean that participants are producing the same hierarchical structure as the DSM. One method of determining the degree of similarity between participants' sortings and the DSM is to calculate a Pearson correlation coefficient between the distance matrices of the two. The correlation coefficient assesses the degree of linear relationship between two variables. If there is a positive correlation, then high values on one variable are associated with high values in the other and vice versa with low values. Calculating these correlations has no bearing on the structure of the data found, only that they preserve the same rank ordering of pairs of disorders. A high correlation with the DSM may exist in data that is not hierarchically structured and a low correlation may exist in data that is hierarchical.

Correlations between participants and the DSM were calculated by concatenating the distance matrices used into a single column vector. Again, there was a high amount of missing data which was excluded pairwise, so individual correlations are based on substantially different n 's. The correlation coefficients and associated n 's between the Georgia, Alabama, and Auburn sample participants and the DSM are shown in Tables 4, 5 and 6, respectively.

For the Georgia data, all correlations with the DSM were positive and significant. However, only two were above .4. The majority of the correlations were in the .2 range, indicating a weak relationship. For the Alabama data, two correlations were not significant with the DSM, and represent true zero order relationships, as the high n greatly inflates the chance of finding a significant

relationship. Of the remaining correlations which were significant, only one was higher than .4, with the majority falling between .1 and .2. Again, there seems to be some degree of relationship among participants to the DSM, but that relationship is weak at best. The correlations between the Auburn participants and the DSM were all significant. Of the 13 participants included in the analysis, only three had correlations lower than .3. The majority of correlations were between .3 and .4, with two correlations reaching as high as .6. Therefore, the novice sample seems to evidence a greater degree of similarity to the DSM than either of the expert samples.

There were a subset of participants in each sample that were highly correlated with each other. It may be that these correlations reflect similar groupings of disorders between participants, which further reflect similar conceptual understandings of the disorders. However, these similarities may be artifacts of missing data. In other words, the high relationships found between some subjects may be a result of having few disorders to compare, where those with higher n 's have more chance to disagree. To test this, I formed three groups of ten correlations each: those that involved two participants that were highly correlated, those that involved two participants who did not correlate with any other participants, and a combination of one correlating participant and one non-correlating participant. The number of stimulus pairs of these three groups were not significantly different; $F(2, 27) = 0.591, p = .561$. Therefore, the correlations found between participants do not seem to be an effect of unequal n 's. The similarity of their structures can be assumed to reflect some similar conception of

disorders across participants. However, an exploration of these structures is beyond the scope of this paper.

Discussion

The results of this study raise a variety of interesting issues. First, and perhaps foremost, participants' organizations of mental disorders are hierarchical. Contrary to the original hypothesis of the study that participants would exhibit no definite structure, participants demonstrated hierarchical structuring to their sortings of disorders across level of expertise and methodology used. Such a finding is consistent with Berlin's (1992) hypothesis that all folk taxonomies have an inherent hierarchical organization. If the organization of mental disorders is a folk taxonomy, it seems to fit with Berlin's general description of other folk taxonomies.

However, this hierarchy exhibited some interesting features. While it is uniform across groups of participants, the hierarchy seen in the sortings is not the same as the "ideal" hierarchy found in the DSM. The hierarchical structure of the DSM can be considered the consensus of a variety of professionals. However, this consensus is not a blind consensus, but is based upon an extensive review of the scientific literature of mental disorders (Widiger et al., 1994, 1996, 1997, 1998). The DSM structure is as close to a standard scientific taxonomy as exists in clinical psychology and psychiatry. Clinicians in the samples studied herein seem to have represented some of the structure seen in the DSM, but they did not replicate that structure. There may be many reasons for the disparity. I offer three explanations that point to broader, epistemic issues

regarding the nature of mental disorders that may play a factor in the sortings of clinicians, as well as one mundane methodological issue that also may have shaped clinicians' sortings.

First, students may leave graduate school with an incomplete understanding of the DSM, and a complete understanding of the DSM is not learned through personal experience. Here, the difference between clinicians and the DSM could be seen as a deficit in learning the DSM. This view assumes that the DSM is the "correct" representation of the state of nature of mental disorders. There is some support for the idea that clinicians have not completely learned the DSM, for in one of very few such surveys, Dempster (1990) found that practicing mental health professionals (not just psychologists) mostly (57.7%) felt that their training in DSM diagnostic categories was inadequate for them to competently perform their jobs. While the terminology of the survey may be ambiguous, the finding does point to the idea that clinicians may not receive adequate training in the DSM, thus the difference in clinicians from the DSM may be associated with a lack of appropriate training in the use of the DSM.

Alternatively, students may leave graduate school with an understanding of the DSM, but that understanding becomes tempered through experience to blur some of the distinctions made between disorders and their relations to other disorders. In other words, clinicians may leave graduate school adequately trained in the DSM, but the actual state of affairs in the world is more ambiguous, and they respond to that ambiguity by adjusting their organization of mental disorders accordingly. Clinicians, in interacting with the actual presentation of

mental disorders in the world, may drift from the “pure” presentation of disorders in the DSM. Pica (1998) suggested that the nature of mental disorders is ambiguous and unclear, and so the ambiguous nature of most graduate training prepares students in kind. Novices trained in one context only seem to be able to transfer that knowledge to new contexts if it was learned in an abstract manner to begin with (Donnelly & McDaniel, 1993; Hinds, Patterson, & Pfeffer, 2001). Therefore, the DSM may not adequately capture the ambiguous nature of mental disorders, and so the clinicians in this study were not expressing an inadequate knowledge of the DSM, but their sortings reflect their individual understanding of the ambiguous state of affairs.

Third, both the scientific study of taxonomy and clinical experience may offer incomplete, distorted views of the “true” state of nature as if looking through a tinted window. The first explanation offered above stated that the DSM is essentially correct, and clinicians are wrong for being different from it. The second stated that the DSM was wrong for not reflecting the true nature of disorders in the world, and clinicians have captured the truth through their experience. A more likely explanation is that something of each is occurring. Mental disorders are by nature ambiguous, fluid, and do not fit simple, essentialist definitions, and so the taxonomy built to describe them reflects that ambiguity. Clinicians in this study, then, are simply doing their best to organize a recognized chaos, and use the tools of taxonomy and experience to guide them. Their difference from the DSM does not reflect any deficiency in either the DSM or their experience, but simply reflects the uncertainty inherent in both. If this is

true, it only makes the consensus found between clinicians all the more remarkable.

Fourth, there is a more mundane possible explanation for why clinicians differ from the DSM that does not include any ontological or epistemological implications. Clinicians in all samples were instructed to sort disorders based upon their treatment, not upon the DSM. The DSM is not a treatment based classification system, but rather tries to remain atheoretical with respect to treatment options (APA, 1994). That is not to say that the DSM is not intended to be useful for treatment, but that a DSM diagnosis is neutral regarding a number of available treatment options open to the clinician. A clinician's theoretical orientation (and theory of etiology) tends to guide the method of treatment used (i.e. cognitive interventions, behavioral work, interpersonal exploration, etc.). Therefore, this choice of instruction may have led participants not to replicate the DSM, even though they may be capable of doing so under other conditions.

One highly striking feature of the hierarchical structures seen in the sortings of participants is the lack of hierarchical differentiation between levels 3 and 4, which corresponds to the differentiation of Axis I disorders from those of Axis II. Hierarchical organization existed at lower levels. Clinicians meaningfully differentiated subclusters of disorders, such as the grouping of depression and dysthymia, from higher order categories, like the mood disorders (which also includes the bipolar disorders and cyclothymia). However, clinicians did not consider the personality disorders and mental retardation (Axis II) as different from Axis I groups of disorders (mood disorders, psychotic disorders, eating

disorders, etc.). Rather, the distance of the personality disorders and mental retardation from these groups was equal to the distance of these groups within themselves. Therefore, clinicians in these samples did not treat Axis II as a meaningful or useful taxonomic distinction, but conceptualized the personality disorders as they would any other major group of disorders.

The personality disorders were originally separated from other mental disorders in the DSM-III (APA, 1980). The reason for doing so was to increase the frequency with which personality disorders were diagnosed, because in previous editions of the DSM, the inclusion of another diagnosis often excluded the possibility of a personality diagnosis (Frances, 1980). In that sense, a comorbid personality disorder may be ignored in favor of more salient Axis I type pathology. However, including the personality disorders on a separate axis simply so that comorbid personality disorders may be recognized did not seem to be sufficient justification for such a dramatic a shift (Spitzer & Williams, 1983), and thus Kendell (1983) suggested that Axis II disorders be distinguished by their stable and chronic nature. However, recent work has shown that personality disorders are not more chronic or more stable than common Axis I disorders such as depression (Shea & Yen, 2003). Other theorists have argued that the comorbidity between Axis I and Axis II pathology may indicate that the disorders in each should not be kept distinct (Krueger & Tackett, 2003; Widiger, 2003). The participants of this study did not functionally consider personality disorders separate from Axis I disorders. Therefore, the results of this study support the view that Axis II should not be separated from Axis I.

I hypothesized that there would be an effect of methodology upon clinicians' sortings; specifically that sortings made under the progressive grouping method would appear more hierarchical than the dimensionally based tabletop sorting. However, this occurred in only one instance. When the methodologies were compared, participants in each produced a nearly identical pattern. The slopes between the higher levels (2 through 4) were parallel across methodology. However, participants in the progressive grouping method evidenced a more substantial slope between levels 1 and 2. Further, this slope was equal to the ideal DSM slope in the progressive grouping methodology where it was not in the tabletop methodology. Nonetheless, both slopes did evidence hierarchical organization. When the methodologies were compared only within experts or only within novices, all slopes were parallel. This difference between methodologies at the lowest level has no easy explanation. The progressive grouping methodology may be more sensitive, or may provide more easily measured distinctions, at these levels than does the tabletop method. However, a simpler explanation is that the difference between methodologies may simply be sampling bias, as the samples involved were small.

Regardless of this difference, whatever structure clinicians hold seems to overwhelm any potential shaping effects of the methodology used, producing nearly identical structures even in cases of small sample size. Such uniformity suggests that different clinicians hold a similar organizational pattern of mental disorders, and that the pattern does not vary drastically across clinicians (although there is certainly a wide variety in the manner this pattern may be

expressed, as clinicians produced sortings that appeared quite different upon surface examination). This finding is encouraging in the sense that clinicians seem to recognize some similarity in the array of psychopathology they encounter, and that they express this similarity in a uniform pattern.

I further hypothesized that differences would exist between the sortings of experts and novices. This hypothesis had intuitive appeal, as one would expect experience to have some effect upon clinicians' understandings of disorders, i.e. that an experienced clinician would have a "better" understanding of the organization of disorders through further exposure to these disorders. The tree experts in the Medin et al. (1997) study did have more complex and scientifically accurate sortings. However, there were very few differences between experts and novices in the current study, and their similarity is more striking than their difference. Only one difference between experts and novices occurred. The novices in the tabletop methodology had more distance within groups at the highest level of the hierarchy (although this did not result in a different slope between levels 3 and 4). Given the overall pattern seen in the other samples, this difference may disappear if more participants were included. However, there may be a genuine difference. Novices are still immersed in the scholastic setting of graduate school, where presumably a student is present to learn the facts of the field and may take such information at face value. Therefore, novices may be more likely to reconstruct the DSM, as they have just learned that it is the "correct" and accepted organization of mental disorders. Novices in other settings have displayed similar results, in that they perform well using concretely

learned instructions only in situations where those instructions directly apply (Hinds, Patterson, & Pfeffer, 2001). Or, similarly, novices may lack the experience to have constructed their own organization of mental disorders, and therefore rely on the organization given to them in the DSM. As novices mature into experts, they may develop a different understanding of the nature of mental disorders. In so doing, they may shift from holding the personality disorders as a separate entity to conceptualizing them as similar to other Axis I groups of disorders. The nature of how novices use and construct concepts seems to shift from using direct similarity to causal and etiological inference as they gain expertise (Shafto & Coley, 2003). This finding needs to be replicated in further work before any such speculations of potential developmental differences between novices and experts warrant further investigation.

Finally, the overall resemblance of clinicians' sortings to the DSM seems moderate at best, as evidenced in their correlations to the DSM. This finding is not surprising, given that clinicians' sortings are hierarchical but not identical to the DSM. The pattern and level of hierarchy found in participants' sortings suggests that there would be a moderate level of resemblance to the DSM. Nonetheless, novices evidenced a higher degree of similarity to the DSM than did experts. Again, this may be due to novices' closer temporal proximity to graduate school, in which they supposedly first learn the accepted organization found in the DSM. Novices may have relied more heavily upon the structure of the DSM to create their own sortings, either out of having learned the information more recently and thereby having less chance to forget it, or through not yet

having the experience to create their own organization of disorders. While novices displayed a higher degree of similarity to the DSM, that similarity does not imply that they had more hierarchical sortings. As we have seen, the overall sortings of experts and novices were largely similar, except for the one difference discussed above.

Limitations of the current study

To the casual observer, the most obvious limitation of this study is the limited size of each sample. While the sample sizes were small, usually one would expect a small sample size to lead to non-significant findings, i.e. a lack of power. Rather, there were consistent differences despite the small sample sizes. Therefore, rather than a limitation, the relation of sample size to the power of the analyses is an indication that the differences observed must be easily observable.

However, the selection of participants and formation of groups may limit the generalizability of the study. The samples were selected as samples of convenience. Assuming that the population of interest in the study is all clinicians, no steps were taken to examine if the samples represent all clinicians in any meaningful way. The participants were almost exclusively psychologists (or psychologists in training), while clinicians can come from a variety of fields, including medicine, social work, school counseling, religious affiliation, etc. Therefore, the results of this study can be generalized only to other psychologists. Further, one could argue that even then, clinicians sampled from regional meetings in Alabama and Georgia may not represent all psychologists

across the nation. However, this statement implies that there are systematic geographical differences in psychologists training or abilities. While there may be such differences, the literature has not addressed this issue. However, there are regional differences in the base rates of various mental disorders (Robins et al., 1984), so the experience of clinicians in different regions must be different. Clinicians in one area may be exposed to more depression than clinicians in another, and thus this differential exposure could lead to different conceptions and organizations of mental disorders. Further, in the expert groups, methodology was confounded with sample. This may have led to systematic differences, but there were no differences in the results for these two groups.

The greatest problem of the study was the large amount of missing data, as participants were allowed to discard unfamiliar diagnoses. The missing data not only were difficult to handle in the initial structuring of the data, but resulted in means for each participant that were based on a largely differing number of stimuli. There were several ways to handle the missing data, and in pilot work, leaving the data as missing had the least extraneous impact on the results. For example, an alternative way to handle the missing data was to impute a distance farther than any other distance in the matrix (i.e. to set all missing values to a value of 8). However, in sortings with large amounts of missing data, that amount of large distances in the matrix pulled the value of the mean distances between groups to a much higher value. Also, having a large portion of the data set to a specific value pulled the correlations between subjects and the DSM close to zero, even though the pattern of the real data points suggest a

relationship. Nonetheless, handling the missing data in the way I did shaped the nature of the results. A different strategy for handling missing data would likely lead to different results and different interpretations. In future studies, it would be useful to require all participants to use every stimulus diagnosis, especially if there were some other way to represent unfamiliarity with the diagnosis.

Future directions

The pattern of results in this study is remarkably striking, given the numerous potential factors that could have driven the results in differing directions. Novices could have been different from experts. Tabletop sortings could have looked qualitatively different than progressive grouping sortings. Samples in differing geographical areas with supposedly differing theoretical orientations and educational histories could have produced systematic differences. However, none of these happened. Instead a uniform pattern of organization emerged despite these factors. This is consistent with Berlin's (1992) conceptualization of the universality of organization within folk taxonomies. Nonetheless, further work needs to be done to confirm the existence of this pattern as an explanatory framework for clinicians' conceptualization of the organization of mental disorders. Only clinical psychologists were tested in this study. Differences may exist across professions, i.e. psychiatrists may hold a different pattern of organization from clinical psychologists, from social workers, etc. Also, the sample sizes of the current study were limited. This always calls into the question the confidence that can be placed on any inferences drawn from such a study. Further testing

across more participants will assuage any such hesitations. The uniformity of the pattern seen in this study begs further exploration.

Also, while a similar pattern was seen, there was no exploration of the factors or structures that may have created this pattern. Clinicians in all these samples may have agreed upon the organization of a subset of the disorders and preserved these groups above all others, while other disorders were organized in idiosyncratic ways and introduce noise into the pattern. The subset of participants in each sample that correlated highly with each other may provide a window into this common structure. While the current study did not address these issues, this is a rich area for further exploration and may illuminate the striking pattern of results seen here. One could potentially examine the frequency with which clinicians preserve select groups seen in the DSM, or one could evaluate the mean distances of such groups to determine if any particular groups are somehow more distinct or cohesive and thereby carry the overall fit of clinicians' sortings.

References

- American Psychiatric Association. (1980). *Diagnostic and statistical manual of mental disorders* (3rd ed.). Washington, DC: Author.
- American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: Author.
- American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders* (4th ed., text revision). Washington, DC: Author.
- Atran, S. (1995). Classifying nature across cultures. In D. N. Osherson & E. E. Smith (Eds.), *An invitation to cognitive science, Volume 3: Thinking* (pp. 131-174). Cambridge, MA: MIT Press.
- Bailenson, J., Shum, M., Atran, S., Medin, D., & Coley, J. (2002). A bird's eye view: Biological categorization and reasoning within and across cultures. *Cognition*, 84, 1-53.
- Berlin, B. (1992). *Ethnobiological classification: Principles of categorization of plants and animals in traditional societies*. Princeton, NJ: Princeton University Press.
- Blashfield, R. K. (2001). The Vulgate DSM-IV: A Review of Am I OK? A Layman's Guide to the Psychiatrist's Bible. *Journal of Nervous & Mental Disease*, 189, 3-7.
- Brown, T.; Chorpita, B.; & Barlow, D. (1998). Structural relationships among dimensions of the DSM-IV anxiety and mood disorders and dimensions of

- negative affect, positive affect, and autonomic arousal. *Journal of Abnormal Psychology, 107*, 179-192.
- Clark, L., & Watson, D. (1991). Tripartite model of anxiety and depression: Psychometric evidence and taxonomic implications. *Journal of Abnormal Psychology, 100*, 313-336.
- Costa, P.T., & Widiger, T.A. (Eds.) (1994). *Personality disorders and the five-factor model of personality*. Washington: American Psychological Association.
- Dempster, L. (1990). How mental health professionals view their graduate training. *Journal of Training and Practice in Professional Psychology, 4*, 4-19.
- Donnelly, C., & McDaniel, M. (1993). Use of analogy in learning scientific concepts. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19*, 975-987.
- Flanagan, E. (2003). *Novice and expert psychologists' natural taxonomies of mental disorders*. Unpublished doctoral dissertation, Auburn University.
- Flanagan, E., & Blashfield, R. (2000). Essentialism and a folk-taxonomic approach to the classification of psychopathology. *Philosophy, Psychiatry, & Psychology, 7*, 183-189.
- Frances, A. J. (1980). The DSM-III personality disorders section: A commentary. *American Journal of Psychiatry, 137*, 1050-1054.
- Frances, A., & First, M. (1998). *Am I Okay?: A layman's guide to the psychiatrist's bible*. New York: Simon and Shuster.

- Hall, W. (1996). What have population surveys revealed about substance use disorders and their co-morbidity with other mental disorders? *Drug and Alcohol Review, 15*, 157-170.
- Hinds, P., Patterson, M., & Pfeffer, J. (2001). Bothered by abstraction: The effect of expertise on knowledge transfer and subsequent novice performance. *Journal of Applied Psychology, 86*, 1232-1243.
- Johnson, K., & Mervis, C. (1997). Effects of varying levels of expertise on the basic level of categorization. *Journal of Experimental Psychology: General, 126*, 248-277.
- Kendell, R. E. (1983). DSM-III: A major advance in psychiatric nosology. In R. L. Spitzer, J. B. W. Williams, & A. E. Skodol (Eds.). *International Perspectives on DSM-III* (pp. 55-68). Washington, DC: American Psychiatric Press.
- Kessler, R. (1995). The National Comorbidity Survey: Preliminary results and future directions. *International Journal of Methods in Psychiatric Research, 5*, 139-151.
- Krueger, R., & Tackett, J. (2003). Personality and psychopathology: Working toward the bigger picture. *Journal of Personality Disorders, 17*, 109-128.
- Langenbucher, J.; Martin, C.; Labouvie, E.; Sanjuan, P.; Bavly, L.; & Pollock, N. (2000). Toward the DSM-V: The withdrawal-gate model versus the DSM-IV in the diagnosis of alcohol abuse and dependence. *Journal of Consulting and Clinical Psychology, 68*, 799-809.

- Lopez, A., Atran, S., Coley, J., Medin, D., & Smith, E. (1997). The tree of life: Universal and cultural features of folkbiological taxonomies and inductions. *Cognitive Psychology*, *32*, 251-295.
- Lynam, D., & Widiger, T. (2001). Using the five-factor model to represent the DSM-IV personality disorders: An expert consensus approach. *Journal of Abnormal Psychology*, *110*, 401-412.
- Medin, D., Lynch, E., Coley, J., & Atran, S. (1997). Categorization and reasoning among tree experts: Do all roads lead to Rome? *Cognitive Psychology*, *32*, 49-96.
- Miller, J.; Reynolds, S.; & Pilkonis, P. (2004). The validity of the five-factor model prototypes for personality disorders in two clinical samples. *Psychological Assessment*, *16*, 310-322.
- Pica, M. (1998). The ambiguous nature of clinical training and its impact on the development of student clinicians. *Psychotherapy*, *35*, 361-365.
- Robins, L. N.; Helzer, J. E.; Weissman, M. M.; Orvaschel, H.; Gruenberg, E.; Burke, J. D.; & Regier, D. A. (1984). Lifetime prevalence of specific psychiatric disorders in three sites. *Archives of General Psychiatry*, *41*, 949-958.
- Romney, A., Weller, S., & Batchelder, W. (1986). Culture as consensus: A theory of culture and informant accuracy. *American Anthropologist*, *88*, 313-338.
- Rosch, E. (1978). Principles of categorization. In E. Margolis & S. Laurence (Eds), *Concepts: Core Readings* (pp 189-206). Cambridge, MA: MIT Press.

- Shafiq, P., & Coley, J. (2003). Development of categorization and reasoning in the natural world: Novices to experts, naïve similarity to ecological knowledge. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 29, 641-649.
- Shea, M. T., & Yen, S. (2003). Stability as a distinction between Axis I and Axis II disorders. *Journal of Personality Disorders*, 17, 373-386.
- Spitzer, R., & Williams, J. (1983). International perspectives: Summary and commentary. In R. L. Spitzer, J. B. W. Williams, & A. E. Skodol (Eds.). *International Perspectives on DSM-III* (pp. 339-353). Washington, DC: American Psychiatric Press.
- Widiger, T. A. (2003). Personality disorder and Axis I psychopathology: The problematic boundary of Axis I and Axis II. *Journal of Personality Disorders*, 17, 90-108.
- Widiger, Thomas A; Frances, Allen J; Pincus, Harold Alan; First, Michael B; Ross, Ruth; & Davis, Wendy (Eds). (1994). *DSM-IV sourcebook, Vol. 1*. Washington, DC: American Psychiatric Publishing, Inc.
- Widiger, Thomas A; Frances, Allen J; Pincus, Harold Alan; Ross, Ruth; First, Michael B; & Davis, Wendy Wakefield (Eds). (1996). *DSM-IV sourcebook, Vol. 2*. Washington, DC: American Psychiatric Publishing, Inc.
- Widiger, Thomas A; Frances, Allen J; Pincus, Harold Alan; Ross, Ruth; et al. (Eds). (1997). *DSM-IV sourcebook, Vol. 3*. Washington, DC: American Psychiatric Publishing, Inc.

- Widiger, Thomas A; Frances, Allen J; Pincus, Harold Alan; Ross, Ruth; First, Michael B; Davis, Wendy; & Kline, Myriam (Eds). (1998). *DSM-IV sourcebook, Vol. 4*. Washington, DC: American Psychiatric Publishing, Inc.
- Zinbarg, R., & Barlow, D. (1996). Structure of anxiety and anxiety disorders: A hierarchical model. *Journal of Abnormal Psychology, 105*, 181-193.

Appendix A

Appendix A displays the set of 67 mental disorders used as stimuli by

Flanagan (2003).

1. depression
2. dysthymia
3. bipolar I
4. bipolar II
5. cyclothymia
6. panic disorder
7. phobias
8. GAD (Generalized Anxiety Disorder)
9. OCD (Obsessive Compulsive Disorder)
10. PTSD (Posttraumatic Stress Disorder)
11. substance dependence
12. substance abuse
13. substance induced disorder
14. anorexia nervosa
15. bulimia nervosa
16. sexual dysfunction
17. paraphilia
18. GID (Gender Identity Disorder)
19. substance-induced sleep disorder
20. sleep disorder due to a medical condition
21. circadian rhythm sleep disorder
22. insomnia
23. hypersomnia
24. nightmare disorder
25. sleep terror disorder
26. sleepwalking disorder
27. OCPD (Obsessive Compulsive Personality Disorder)
28. dependent
29. avoidant
30. histrionic
31. narcissistic
32. borderline
33. paranoid
34. antisocial
35. schizoid

36. schizotypal
37. focus on physical symptoms (somatization disorder, conversion disorder)
38. focus on fear of disease (hypochondriasis, body dysmorphic disorder)
39. intermittent explosive disorder
40. kleptomania
41. pyromania
42. pathological gambling
43. trichotillomania
44. dissociative amnesia
45. MPD (multiple personality disorder/ dissociative identity disorder)
46. depersonalization disorder
47. adjustment disorder
48. substance-induced psychotic disorder
49. psychotic disorder due to a medical condition
50. depression or mania with psychotic features
51. brief psychotic disorder
52. schizophrenia
53. delusional disorder
54. schizoaffective disorder
55. shared psychotic disorder
56. delirium
57. dementia
58. amnesia
59. mental retardation
60. autism
61. conduct disorder
62. ODD (oppositional-defiant disorder)
63. ADHD (attention deficit/hyperactivity disorder)
64. tic disorder
65. encopresis
66. enuresis
67. separation anxiety disorder

Appendix B

Appendix B displays the hierarchical organization of the DSM-IV (1994) as implied by its table of contents.

I. Axis I disorders

A. childhood disorders

1. Autism
2. disruptive behavior disorders
 - a. Attention-Deficit/Hyperactivity Disorder (ADHD)
 - b. Conduct Disorder
 - c. Oppositional Defiant Disorder (ODD)
3. tic disorder
4. elimination disorders
 - a. Encopresis
 - b. Enuresis
5. Separation Anxiety Disorder

B. cognitive disorders

1. Delerium
2. Dementia
3. Amnesia

C. substance-related disorders

1. substance dependence
2. substance abuse

D. psychotic disorders

1. Schizophrenia
2. Schizoaffective Disorder
3. Delusional Disorder
4. Brief Psychotic Disorder
5. Shared Psychotic Disorder
6. Substance Induced Psychotic Disorder
7. Psychotic Disorder due to a Medical Condition

E. mood disorders

1. depressive disorders
 - a. Depression
 - b. Dysthymia
2. bipolar disorders
 - a. Bipolar I Disorder

- b. Bipolar II Disorder
 - c. Cyclothymia
- F. anxiety disorders
 - 1. Panic Disorder
 - 2. Phobias
 - 3. Obsessive Compulsive Disorder (OCD)
 - 4. Posttraumatic Stress Disorder (PTSD)
 - 5. Generalized Anxiety Disorder (GAD)
- G. somatoform disorders
 - 1. focus of physical symptoms (e.g. somatization disorder)
 - 2. focus on fear of disease (e.g. hypochondriasis)
- H. dissociative disorders
 - 1. Dissociative Amnesia
 - 2. Multiple Personality Disorder (MPD)
 - 3. Depersonalization Disorder
- I. sexual disorders
 - 1. Sexual Dysfunction
 - 2. Paraphilia
 - 3. Gender Identity Disorder
- J. eating disorders
 - 1. Anorexia
 - 2. Bulimia
- K. sleep disorders
 - 1. dyssomnias
 - a. Insomnia
 - b. Hypersomnia
 - c. Circadian Rhythm Sleep Disorder
 - 2. parasomnias
 - a. Nightmare Disorder
 - b. Sleep Terror Disorder
 - c. Sleepwalking
 - 3. other sleep disorders
 - a. Substance-Induced Sleep Disorder
 - b. Sleep Disorder due to a Medical Condition
- L. impulse-control disorders
 - 1. Explosive Disorder
 - 2. Kleptomania
 - 3. Pyromania
 - 4. Pathological Gambling
 - 5. Trichotillomania
- M. Adjustment Disorder

- II. Axis II Personality Disorders
 - A. Cluster A personality disorders
 - 1. Paranoid

- 2. Schizoid
- 3. Schizotypal
- B. Cluster B personality disorders
 - 1. Antisocial
 - 2. Borderline
 - 3. Histrionic
 - 4. Narcissistic
- C. Cluster C personality disorders
 - 1. Avoidant
 - 2. Dependent
 - 3. Obsessive Compulsive Personality Disorder (OCPD)
- D. Mental Retardation

Note: Frances and First (1998) include Depression or Mania with Psychotic Features as a separate disorder, and this category is included within the 67 stimuli used. However, this diagnosis is not recognized in the DSM-IV, and can conceivably be included with the psychotic or mood disorders.

Table 1

Means and standard deviations across levels of the hierarchy for the combined sample by methodology

	Level 1	Level 2	Level 3	Level 4
Progressive Grouping				
<i>M</i>	4.44	5.36	6.41	6.49
<i>SD</i>	0.73	0.70	0.36	0.42
Tabletop				
<i>M</i>	1.04	1.45	2.51	2.74
<i>SD</i>	0.51	0.49	0.43	0.57

Table 2

Means and standard deviations across levels of the hierarchy for the expert samples by methodology

	Level 1	Level 2	Level 3	Level 4
Progressive Grouping (GA)				
<i>M</i>	4.49	5.09	6.31	6.39
<i>SD</i>	0.66	0.64	0.40	0.52
Tabletop (AL)				
<i>M</i>	1.09	1.47	2.45	2.58
<i>SD</i>	0.50	0.40	0.34	0.46

Table 3

Means and standard deviations across levels of the hierarchy for the novice (Auburn) sample by methodology

	Level 1	Level 2	Level 3	Level 4
Progressive Grouping				
<i>M</i>	4.39	5.60	6.50	6.57
<i>SD</i>	0.83	0.69	0.32	0.33
Tabletop				
<i>M</i>	0.85	1.37	2.75	3.40
<i>SD</i>	0.57	0.81	0.70	0.57

Table 4

Correlations of the sortings of the Georgia sample participants with the DSM

Participant	<i>r</i>	<i>n</i>
1	.326	666
2	.458	378
3	.527	1326
4	.237	630
5	.262	1431
6	.278	1176
7	.260	1176

Note: All correlations have a $p < .001$.

Table 5

Correlations of the sortings of the Alabama sample participants with the DSM

Participant	<i>r</i>	<i>n</i>
1	.187*	351
2	.198*	861
3	.261*	1540
4	.270*	1326
5	.177*	1176
6	.096	325
7	.487*	903
8	.217*	595
9	.284*	1485
10	.325*	703
11	.052	903
12	.309*	1431
13	.102*	1540
14	.114*	861
15	.095*	1275
16	.207*	2145
17	.184*	1035
18	.338*	595
19	.219*	253
20	.193*	1275
21	.109*	780

* These correlations have a $p \leq .001$; the others are non-significant.

Table 6

Correlations of the sortings of the Auburn sample participants with the DSM

Participant	<i>r</i>	<i>n</i>
1	.153*	231
2	.410**	820
3	.603**	561
4	.327**	190
5	.331**	171
6	.258**	351
7	.457**	378
8	.353**	406
9	.386**	378
10	.390**	561
11	.225*	136
12	.326**	190
13	.646**	171

* $p < .05$

** $p < .001$

Figure Captions

Figure 1. Mean distances within groups and 95% confidence intervals across levels of the hierarchy for the combined samples by methodology

Figure 2. Mean distances within groups and 95% confidence intervals across levels of the hierarchy for the expert samples by methodology

Figure 3. Mean distances within groups and 95% confidence intervals across levels of the hierarchy for the novice sample by methodology

Figure 4. Comparison of the mean distances within groups and 95% confidence intervals across expert status and methodology

Figure 1

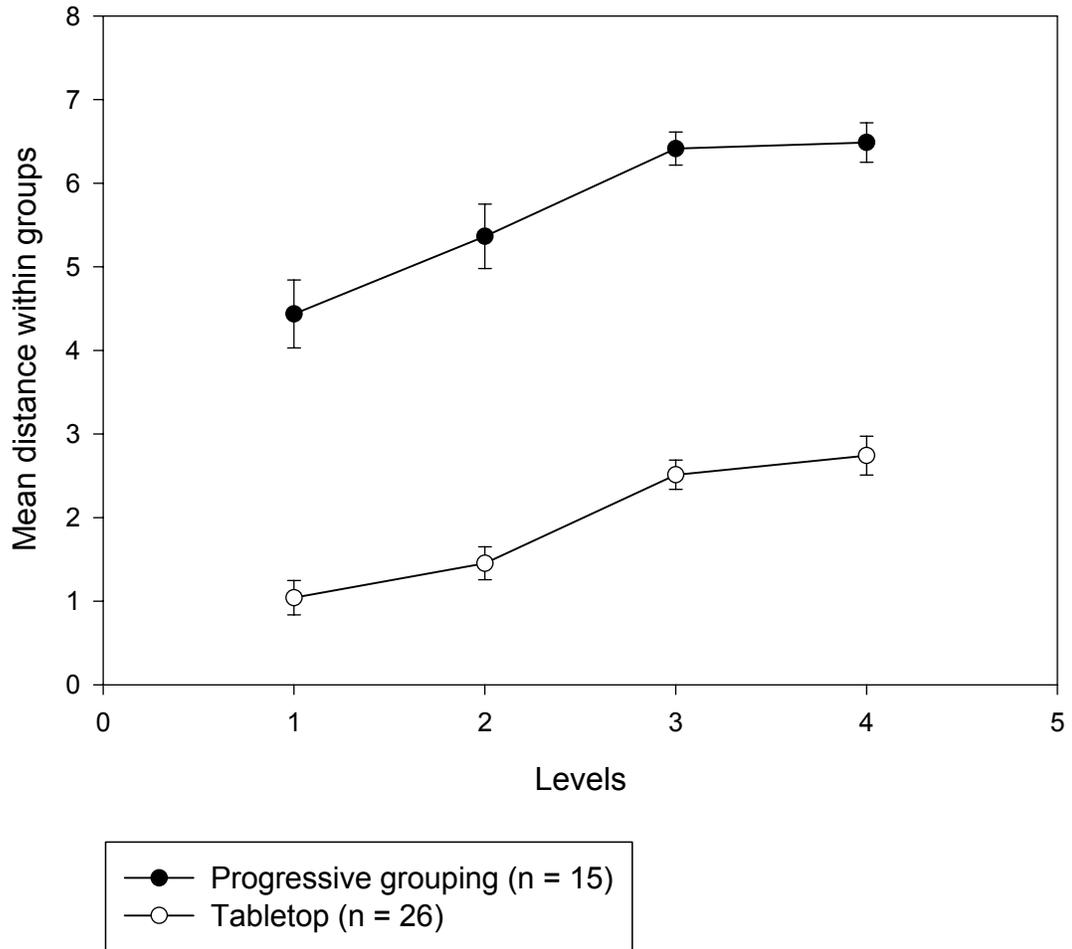


Figure 2

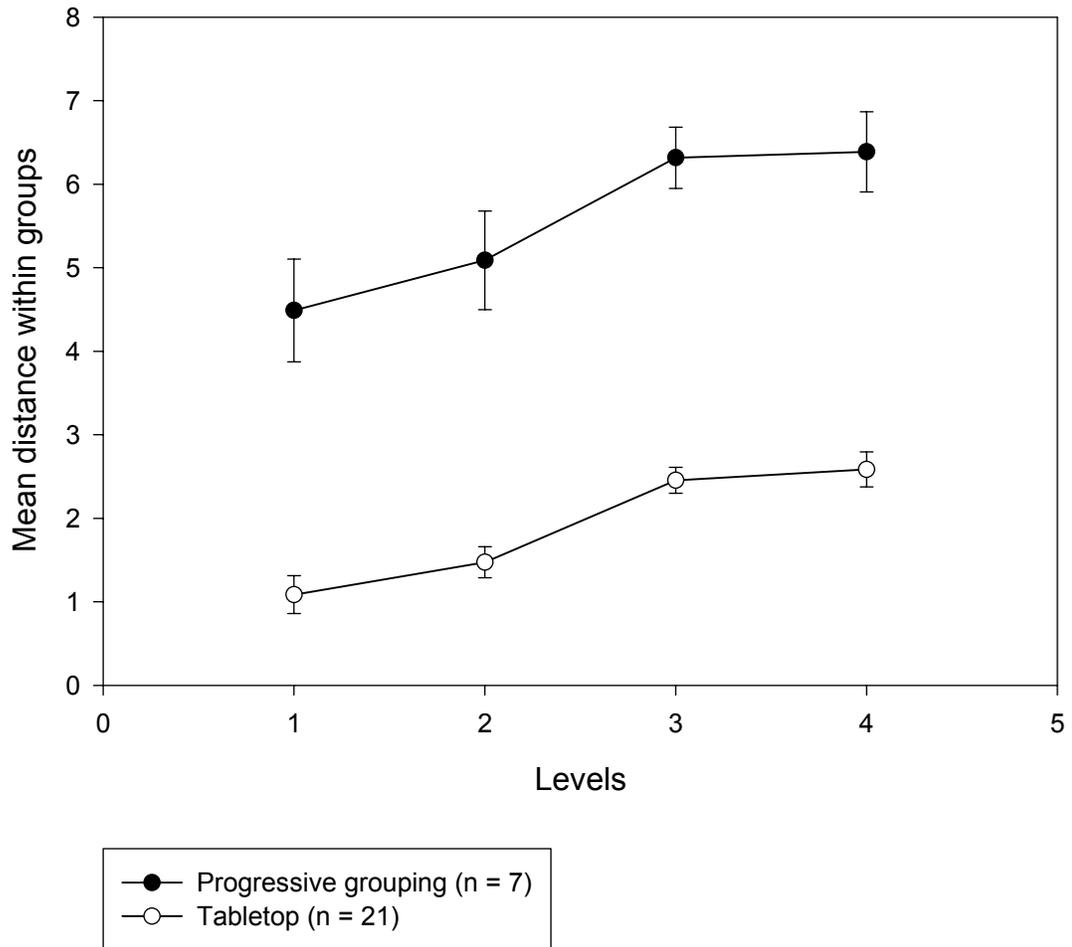


Figure 3

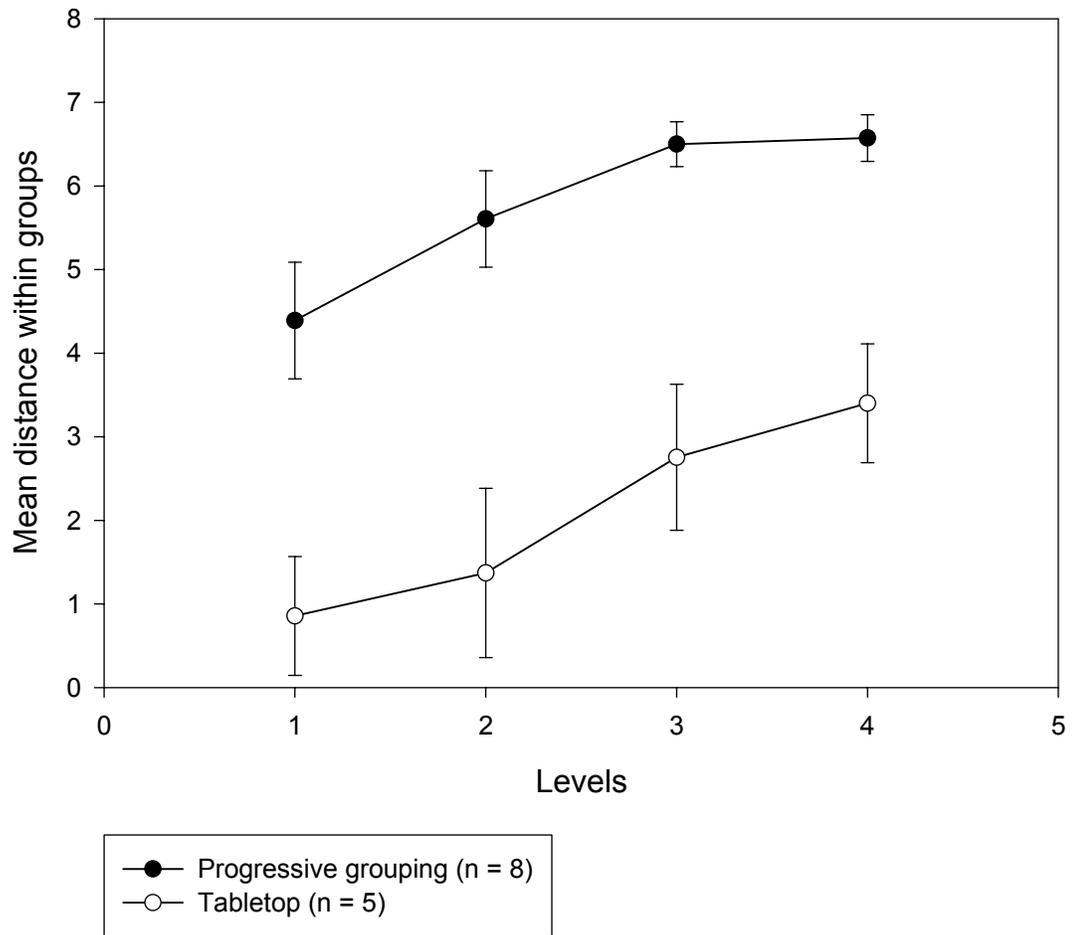


Figure 4

