COMPARING THE UTILITY AND RELIABILITY OF
TWO CURRENT SUICIDE-RELATED NOMENCLATURES

A Dissertation
Presented to
The Graduate Faculty of The University of Akron

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Thomas J. Rankin
May, 2013
COMPARING THE UTILITY AND RELIABILITY OF
TWO CURRENT SUICIDE-RELATED NOMENCLATURES

Thomas J. Rankin
Dissertation

Approved:

Advisor
Dr. Ronald Levant

Accepted:

Department Chair
Dr. Karin Jordan

Committee Member
Dr. Philip Allen

Associate Dean of the College
Dr. Evonn Welton

Committee Member
Dr. Dawn Johnson

Dean of the Graduate School
Dr. George Newkome

Committee Member
Dr. John Queener

Date

Committee Member
Dr. Ingrid Weigold
This dissertation examined the reliability and perceived clinical utility of two current suicide-related nomenclatures. The language of suicidology has been confusing for the past half-century, thereby inhibiting effective study of the problem of suicide. This dissertation’s survey research compared two nomenclatures: the Columbia Classification Algorithm of Suicide Assessment (C-CASA; Posner, Oquendo, Gould, Stanley, & Davies, 2007) and the Silverman nomenclature (Silverman, Berman, Sanddal, O’Carroll, & Joiner, 2007a and 2007b). Hypothetical vignettes were constructed to represent the different terms and definitions of each of the two nomenclatures; ten vignettes were randomly selected to be presented to participants. The 131 participants were mental health clinicians who were currently seeing clients for at least five hours per week. The hypotheses were that the C-CASA would exhibit better reliability and significantly better perceived clinical utility than the Silverman nomenclature. Perceived clinical utility was measured by a new 13-item questionnaire constructed for the purposes of this study. An exploratory factor analysis indicated that the perceived clinical utility questionnaire had a stable one-factor structure and an alpha of 0.95. The data supported both hypotheses. C-CASA’s reliability, as measured by kappa statistics, was 0.59, while the Silverman nomenclature’s was 0.45; this was more than a 0.1 difference in C-CASA’s favor. On the 7-point Likert-type scale of the perceived clinical utility questionnaire, the Silverman nomenclature received positive scores that were a bit above neutral (4.62), while the C-CASA received scores almost a point higher (5.43), which were almost halfway between “neutral” (4) and “strongly agree” (7). These differences were statistically significant and represented a large effect size (Cohen’s d was 0.83). These differences suggested that the C-CASA was preferable to the Silverman nomenclature, both in terms of reliability and perceived clinical utility. If replicated by other researchers, this result would mean that the C-CASA showed more promise than the Silverman nomenclature for constructing a common language for suicide-related thoughts and behaviors among practicing clinicians.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>LIST OF TABLES</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>LIST OF FIGURES</td>
<td>viii</td>
</tr>
</tbody>
</table>

## CHAPTER

### I. STATEMENT OF THE PROBLEM

- Comparing Two Major Current Suicide Nomenclatures
- Suicide Nomenclature and Its Relationship to Suicide Rates
- Absence of a Common Language: A Major Problem for the Study of Suicide
- The Relationship between Suicide Nomenclature and Classification
- The Importance of Testing Suicide-Related Nomenclatures through Reproducible Research
- The DSM-5 Field Trial Protocol Provides a Blueprint for Studying Suicide-Related Nomenclatures
- The Meaning of Reliability for Suicide-Related Nomenclatures
- Shared and Unique Aspects of the Two Nomenclatures
- The Distinction between Validity and Reliability for Suicide-Related Nomenclatures
- The Distinction between the Validity and the Clinical Utility of a Nomenclature
- Summary of the Research Questions

### II. LITERATURE REVIEW

- Introduction
- Literature Review of the Columbia Classification Algorithm of Suicide Assessment (C-CASA)
- Literature Review of the Silverman Nomenclature
- Collecting Evidence Regarding Suicide-Related Nomenclatures
# LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Sample table to assist in understanding kappa</td>
<td>66</td>
</tr>
<tr>
<td>2</td>
<td>Overall kappas, individual vignette kappas, and percent agreement statistics</td>
<td>84</td>
</tr>
<tr>
<td>3</td>
<td>Perceived clinical utility questionnaire factor and loadings from Principal</td>
<td>88</td>
</tr>
<tr>
<td></td>
<td>Axis Factoring (pattern matrix) for the one-factor solution</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>ANOVA statistics for the one-factor solution's mean scores on perceived clinical utility</td>
<td>90</td>
</tr>
<tr>
<td>5</td>
<td>Descriptive statistics for scores on the perceived clinical utility questionnaire for both one-factor and two-factor solutions</td>
<td>91</td>
</tr>
</tbody>
</table>
## LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Scree plot for exploratory factor analysis of the perceived clinical utility questionnaire</td>
<td>87</td>
</tr>
</tbody>
</table>
CHAPTER I

STATEMENT OF THE PROBLEM

Comparing Two Major Current Suicide Nomenclatures

This dissertation will compare evidence regarding the reliability and utility of two major current suicide-related nomenclatures. The first major suicide-related nomenclature was laid out in two successive articles by Silverman, Berman, Sanddal, O’Carroll, and Joiner (2007a & 2007b) and will be called “the Silverman nomenclature.” It was a revision of a groundbreaking suicide nomenclature created by O’Carroll et al. (1996; “the O’Carroll nomenclature”) in their seminal article, “Beyond the Tower of Babel: A Nomenclature for Suicidology”; both Silverman and O’Carroll were co-authors of the 1996 and 2007 articles. The second major suicide-related nomenclature was delineated by Posner, Oquendo, Gould, Stanley, and Davies (2007), and entitled the Columbia Classification Algorithm of Suicide Assessment (“C-CASA”); the C-CASA nomenclature was used to classify suicide-related events in the Federal Drug Administration’s pediatric study of the suicide risk of antidepressants.

The terms and definitions that make up these two nomenclatures are the central focus of this dissertation. It would be premature to present the nomenclatures in their entirety at this early point. However, the terms and definitions of each nomenclature are presented in their entirety in Appendix A. By presenting those terms and definitions in an appendix, the reader can easily locate and review the content of the two nomenclatures at any point while reading this dissertation. It is suggested that the reader examine the terms and definitions of the nomenclatures at some point before reading the second chapter of this dissertation; that way the reader will understand what is being reviewed during the literature review.
These two major suicide nomenclatures each have important conceptual similarities to each other; in fact, one of the leading self-injury and suicide scholars, Harvard professor Matthew Nock, designated the O’Carroll nomenclature, the Silverman nomenclature, and the C-CASA nomenclature as “recent consensus papers” regarding suicide nomenclature (Nock et al., 2008). However, there are important differences between the two nomenclatures: for example, the C-CASA nomenclature includes “preparatory” behaviors, while the Silverman nomenclature does not, probably because preparatory behaviors are “suicidal behaviors… that are exploratory in nature and in need of further research.” (Wenzel, Brown, & Beck, 2009, p. 20). Conversely, the Silverman nomenclature defines “suicide-related communications” such as “suicide threats” and “suicide plans,” whereas the C-CASA does not include those terms as part of its nomenclature.

Sorting out which terms and definitions are supported by evidence is an important priority in suicide research (Silverman et al., 2007a). There is broad consensus among researchers that a good suicide nomenclature is essential to permit accurate classification of suicide-related thoughts and behaviors (Silverman, 2006). Accurate classification facilitates adequate compilation of statistics regarding the public health problem of suicide and attempted suicide (De Leo, Burgis, Bertolote, Kerkhof, & Bille-Brahe, 2006). The scope of the problem of suicide will be sketched below, followed by a summary of how a reliable and clinically useful suicide nomenclature would advance the study of suicidology.

**Suicide Nomenclature and Its Relationship to Suicide Rates**

Suicide is a major public health problem worldwide and accounts for almost 37,000 deaths each year in the United States (CDC, National Centers for Injury Prevention and Control, Web-Based Injury Statistics Query and Reporting System (“WISQARS”), 2012). These statistics are based on the year 2009 national data from the U.S. Centers for Disease Control (CDC), which is the most current year for which statistics are available. In 2009 there were 36,909 deaths by suicide, making suicide the 10th-leading cause of death in the United States; suicide causes more deaths each year in the U.S. than does homicide, chronic liver disease and cirrhosis, hypertension, or Parkinson’s disease (CDC, National Centers for Injury
Prevention and Control, WISQARS, 2012; McIntosh, 2012). The annual U.S. rate of suicide is 12.0 per 100,000 persons, or 1.5% of all U.S. deaths each year, which represents more than 100 deaths in the United States each day by suicide (McIntosh, 2012). Globally, more than one million people die by suicide annually (World Health Organization, 2012a) and the overall rate of suicide is 16 per 100,000 persons (World Health Organization, 2012b).

However, what does it mean to say that the global rate of suicide is 16 per 100,000 persons? The Comprehensive Textbook of Suicidology provides a clear, understandable explanation of suicide rates:

“Suicide rates are normally calculated for each year. For example, a suicide rate of 12 would mean that each year (say, in the United States) for every 100,000 persons in the general population there would be 12 suicides.” (Maris, Berman, & Silverman, 2000, p. 74). Given that the world population exceeded seven billion people as of March 2012 (United States Census Bureau, 2012), the 16 per 100,000 global suicide rate extrapolates to more than 1.1 million persons expected to die by suicide worldwide during the next year.

Suicide deaths, however, are just one category contained in an adequate suicide nomenclature. In fact, choosing the categories contained in a suicide-related nomenclature is one of the most important intellectual decisions involved in constructing a successful nomenclature, since “terminology and definitions are the two elements that constitute a standard nomenclature.” (De Leo et al., 2006, p. 5). The research literature does make clear that certain content areas are important; for example, making sense of what makes death by suicide more or less likely involves tracing the prevalence of suicidal ideation and its relationship to suicide death (Brown, Beck, Steer, & Grisham, 2000). But how do we know whether it is better to label that content by the term “suicidal ideation” (which is used by the C-CASA) or the term “suicide-related ideations” (which is preferred by the Silverman nomenclature)? Likewise, we might conclude from studying the research literature that it is clinically important to include non-suicidal self-injury as a category within a suicide-related nomenclature (Nock, Holmberg, Photos, & Michel, 2007); however, how do we decide whether it is better to use the term “self-harm” (used by the Silverman
nomenclature) or “self-injurious behavior” (used by the C-CASA nomenclature)? Or might the evidence show that those terms are equally reliable and useful to clinicians, such that it would not make a functional difference which term is chosen? In part, these decisions should turn on which terms will be most readily embraced by most clinicians, but it is also important to select terms that make communication and research most clear (Silverman et al., 2007a). Some of the consequences of the absence of a generally accepted suicide nomenclature for clinicians and researchers will be described below.

**Absence of a Common Language: A Major Problem for the Study of Suicide**

Research into suicide has long been seriously hampered by the lack of a consistently used and understood terminology (Maris, 2002). Sixteen years ago, O’Carroll and colleagues described the problem well: “Despite hundreds of years of writing and thinking about suicide… there is to this day no generally accepted nomenclature for referring to suicide-related behaviors – not even at the most basic, conversational level.” (1996, p. 238). In other words, researchers lack a common language to communicate effectively with each other about what they are studying. This is still the case today: De Leo et al. (2006) described existing suicide terminology and definitions as both being “currently confusing.” (p. 5).

O’Carroll et al. (1996) summarized the practical import of this absence of a common language in the field of suicidology:

> If one clinician says to the second, “I admitted a suicide attempter last night,” the second clinician does not know whether that patient was injured in any way, whether the patient actually engaged in any self-harm behavior, or even whether that patient was actually trying to end his or her life. (p. 238)

This ambiguity about whether a patient intended to kill himself or herself applies not only to instances of attempted suicide, but even to some cases in which someone has died (Rosenberg et al., 1988); in other words, because the word “suicide” is understood differently by different people, “suicide” terminology faces some of the same challenges as “attempted suicide” terminology (De Leo et al., 2004).
The absence of a consensus nomenclature also hinders a concerted national effort to understand the confluence of factors that leads an individual to attempt suicide (Silverman et al., 2007a; Mośćicki, 2001). Though the system of certifying suicide deaths is imperfect, there is at least a system in place that tries to track statistics on completed suicide. In contrast, no such system currently exists for tracking attempted suicide (Mośćicki, 2001). In her article on the epidemiology of suicide and attempted suicide, Mośćicki (2001) emphasized that a national surveillance system for systematically collecting data on attempted suicide is essential to understand the phenomenon of attempted suicide, and that this in turn depends upon a well-constructed, standardized nomenclature. "There is a need for an agreed-upon, standardized nomenclature that can be used to reliably and consistently describe attempted suicide events," the absence of which implies that "clinical, public health, and policy decisions will continue to be limited." (Mośćicki, p. 313).

It is not just epidemiology that suffers, however: "One of the biggest difficulties in the reliable diagnosis, assessment, treatment, and prevention of suicide worldwide is the lack of a consistent, common nomenclature and classification procedure for suicidal acts..." (Maris, 2002, p. 319). In other words, clinicians struggling to assist suicidal patients are hamstrung by the absence of a standard suicide-related nomenclature, since the inability to clearly define what a patient struggles with prevents the clinician from accurately classifying the patient’s risk and selecting an appropriate treatment that matches the suicide risk (O’Carroll et al., 1996). To put the point more generally, "classification presupposes an established nomenclature, much in the same way that validity presupposes reliability." (Wenzel et al., 2009, p. 20). If classification systems, such as suicide assessment measures, do not utilize terminology that is well accepted across researchers, then comparisons between studies and measures are difficult to make (Nock et al., 2007). For example, a recent review of adolescent suicidality concluded that “there is marked variability across the studies reviewed in definitions of suicidal ideation and behavior,” and as a result the authors “highly recommended” (King & Merchant, 2008, p. 191) utilizing some consistent nomenclature that would permit meaningful comparison across studies.
The Relationship between Suicide Nomenclature and Classification

Without a reliable nomenclature, we cannot advance to the point at which we can accurately classify suicide-related thoughts and behaviors; appropriate classification presupposes an agreed-upon nomenclature (O’Carroll et al., 1996). The practical problems encountered in classification should inform revisions of terminology, just as classification systems should be founded upon a standard nomenclature (Silverman et al., 2007a). For example, a recent study attempted to conduct a meta-analysis of “suicidal intent” but was unable to do so due to the widely disparate ways that “intent” was operationalized in the 44 studies examined: "Unless terms and variables are clearly defined and uniformly measured, the ability to make meaningful interstudy comparisons is severely limited." (Hasley et. al, 2008, p. 589). Classification systems can be thought of as ways of operationalizing suicide nomenclature, and as such it is just as important to work toward consensus in classification as it is to work toward consensus in nomenclature (Brenner et al., 2011).

The difference between classification and nomenclature was described by O’Carroll et al. (1996). O’Carroll et al. defined nomenclature as “a set of commonly understood, logically defined terms,” (p. 240) which could be used to clearly communicate about differences and similarities observed within the subject matter of the nomenclature: “the terms of any nomenclature may be considered a type of shorthand by which communication about classes of more subtle phenomena is facilitated.” (1996, p. 240). A classification system goes beyond a nomenclature, including the following elements described by O’Carroll et al. (1996):

A systematic arrangement of items in groups or categories, with ordered, nested subcategories; scientific (e.g., biologic or etiologic) validity; exhaustiveness; accuracy sufficient for research or clinical practice; and an unambiguous set of rules for assigning items to a single place in the classification scheme. (p. 240)

As described above, a classification scheme is more complex than a nomenclature, though the terms defined in a nomenclature are essential in making the classification system accurate and thorough in assigning items to a unique place in the classification system.
With this more detailed understanding of the difference between classification systems and nomenclatures, the purpose stated in the first sentence of this chapter can be restated with a slightly different emphasis: “This dissertation will compare evidence regarding the reliability and utility of two major current suicide-related nomenclatures.” The italics highlight that the purpose of this dissertation is to compare and contrast two nomenclatures – not to compare and contrast two classification systems. This is the case even though the C-CASA nomenclature is defined as part of its classification system, whereas the Silverman nomenclature is not contained within a classification system; rather, the Silverman nomenclature was defined in the aforementioned two articles which were devoted specifically to nomenclature, not classification (Silverman 2007a & 2007b).

There are three reasons for focusing on nomenclature rather than classification. First, as was just mentioned, Silverman’s nomenclature is not also a classification system, unlike C-CASA; therefore, it would be impossible to compare the Silverman nomenclature to C-CASA if classification systems were being compared. Second, the classification system involved in C-CASA is more complex than its nomenclature: for example, they involve learning to apply “operationalized guidelines for inference of suicidal intent” for the C-CASA (Posner et al., 2007, p. 1036). By limiting this study to nomenclatures, the project becomes cleaner and more manageable. Third, given that a solid nomenclature is necessary for any classification system to work well, focusing on comparing and contrasting the two nomenclatures should ultimately contribute to future research regarding classification systems (Silverman, 2007b).

The Importance of Testing Suicide-Related Nomenclatures through Reproducible Research

Though suicidologists have spilled much ink about the importance of nomenclature, relatively little research has provided reproducible data that illuminates whether particular nomenclature systems, terms, or definitions were supported by research evidence. As the literature review in the next chapter will summarize, the little research that has been done concerning suicide-related nomenclatures has typically taken the form of summaries of expert workgroups or informal feedback (Silverman, 2007a). While expert workgroups and informal feedback have their place in scientific research, they are limited in
that other researchers cannot independently replicate their results, which is an essential feature of good science (Kendell & Jablensky, 2003).

To give one prominent example of data that is valuable but not reproducible, Silverman et al. (2007a) discussed their decision to change the O’Carroll nomenclature’s term “instrumental suicide-related behavior” to the Silverman nomenclature’s term “suicide-related communication”:

Although we felt that O’Carroll et al.’s term, "instrumental suicide-related behavior" (ISRB), came very close to accurately identifying and defining the self-destructive behavior which is not intended to result in death, yet is potentially self-injurious and containing a communication component, we were persuaded by our colleagues that the term "ISRB" is a "mouthful" which can be easily misunderstood, and would be difficult to insert into the existing vocabulary of suicidology. (p. 257)

Although the decision to change the terminology might well have been the correct one, this methodology represents the dominant paradigm in suicide-related nomenclature construction: suicidologists talk with each other and with clinicians regarding potential or actual problems encountered with specific terminology, and then report their decisions in journal articles. However, there are no reproducible data presented to back up these decisions – instead, we are left trusting the experts to “get it right” with respect to the terms selected and definitions provided.

This absence of reproducible research may be part of the problem that suicidology has had, as a field, in obtaining a widespread, standard, consensus nomenclature that can be used by clinicians and researchers alike. Data supporting the reliability and utility of the terminology and definitions of particular nomenclatures are either not collected, not reported in a way that others can replicate, or both. This is not consistent with the scientific method, which relies on data and replication to advance the state of our knowledge: “It is also a fundamental requirement of all scientific research that the subject matter of the investigation should be described with sufficient accuracy and in sufficient detail to enable others to repeat the study if they wish to." (Kendell & Jablensky, 2003, p. 10). Even in highly conceptual matters such as constructing a suicide-related nomenclature, theoretical conceptual analysis ought to be joined with scientific method in order to produce the most defensible results (Haig, 2011).
Given the lack of prior research collecting reproducible data on suicide-related nomenclature, it was initially difficult to devise a way of empirically assessing the reliability and utility of a suicide-related nomenclature. Within the field of suicidology, there were no examples of prior research methods to guide this project. However, there was a content area related to the study of suicide in which there was already an established protocol for collecting evidence regarding a clinically important nomenclature: research concerning the reliability and clinical utility of the *Diagnostic and Statistical Manual of Mental Disorders* (DSM) of the American Psychiatric Association. The reader is likely aware that the American Psychiatric Association is currently in the process of revising the DSM; the DSM-IV-TR (2000) will be updated to the DSM-5 in May of 2013 (American Psychiatric Association, 2012a). I contend that the research protocol used by the DSM-5 field trials in routine clinical practice settings can be adapted to effectively test the reliability and utility of the two major current suicide-related nomenclatures.

**The DSM-5 Field Trial Protocol Provides a Blueprint for Studying Suicide-Related Nomenclatures**

The DSM-5 field trials protocol for routine clinical practice settings provides an excellent blueprint for empirical research concerning suicide-related nomenclatures. The APA conducted field trials of new proposed diagnostic criteria from 2010 through 2012 (APA, 2012b). The research agenda for the DSM-5 was developed through a series of 10 conferences jointly sponsored by the APA, the National Institutes of Health (NIH), and the World Health Organization (WHO) from 1999 through 2007 (APA, 2012b). The DSM-5 Research Group collaborated to create the “Field Trial Testing in Routine Clinical Practices” research protocol that was used to test the reliability and utility of new proposed diagnostic criteria (APA, 2012c). That field testing protocol can be adapted to fit empirical assessment of suicide-related nomenclatures because the DSM is itself a nomenclature.

One objection to using the DSM-5 field testing protocol to test suicide-related nomenclatures might be that the DSM is not really a nomenclature, but rather a diagnostic manual that functions as a classification system. However, Kendell and Jablensky, in their 2003 *American Journal of Psychiatry* article entitled, “Distinguishing between the validity and utility of psychiatric diagnoses,” refer to the authors of
the DSM-IV as “authors of contemporary nomenclatures,” (p. 5) thereby explicitly labeling the DSM-IV as a nomenclature. While it may be true that the primary use of the DSM by clinicians is as a classification system, the American Psychiatric Association’s (2012) own current online summary description of the DSM makes clear that it also functions as a nomenclature:

DSM is published by the American Psychiatric Association (APA) and contains descriptions, symptoms, and other criteria for diagnosing mental disorders. These criteria for diagnosis provide a common language among clinicians…. By clearly defining the criteria for a mental disorder, DSM helps to ensure that a diagnosis is both accurate and consistent… from one clinician to another, and means the same thing to both of these clinicians. (p. 1)

The APA refers to “defining the criteria for a mental disorder”; definition is one of the two central aspects of a nomenclature, while terminology is the other (O’Carroll et al., 1996; De Leo et al., 2006). The APA’s characterization of those definitions as providing a “common language” highlights that the DSM has effectively created a nomenclature that is in nearly universal use among clinicians in the United States, with the terminology of that common language being the diagnostic categories of the DSM (Sanders, 2011; Kendell & Jablensky, 2003; Robins & Cottler, 2004). Regardless of one’s feelings about the content of the DSM or the history of controversy within psychology about its benefit to clients, there is no doubt that the interweaving of the DSM into U.S. legal and insurance requirements for treatment (Regier, 2003) have made its definitional criteria the common language of today’s mental health professionals (Sanders, 2011).

Furthermore, the history of the DSM demonstrates that the explicit intention of the creators of the DSM-III was to contribute to a common language for mental health clinicians and researchers. Sanders (2011) noted that “the introduction of the DSM-III emphasizes the importance of having a common diagnostic language” (p. 398); in fact, the introduction to the DSM-III stated, “Clinicians and research investigators must have a common language with which to communicate…” (American Psychiatric Association (“APA”), 1980, p. 1). Similarly, the introduction of the DSM-IV describes the manual’s contents as defining diagnoses because they are useful to clinical practice, not because the diagnostic categories exist in some existential sense: “there is no assumption that each category of mental disorder is a
completely discrete entity with absolute boundaries dividing it from other mental disorders or from no mental disorder.” (APA, 1994, p. xxii).

To the extent that people have doubts about the validity of the DSM, this only supports its status as a nomenclature. For example, two authors from the Washinton University School of Medicine – namely Robins and Cottler (2004) – argued that all versions of the DSM since the DSM-III have functioned as nomenclatures precisely because of their diagnoses’ lack of correspondence to biologically-based categories found in the world:

The official nomenclatures from 1980 onward have greatly improved communication... Epidemiologic diagnostic results, for which interviews faithful to the official nomenclature are used, can be correctly understood by anyone consulting the official diagnostic manual. Otherwise, there is room for endless doubts about whether persons given a positive diagnosis “really” had that disorder. Psychiatry does not yet have convincing ways to recognize “real” disorders; till [sic] then, we will have to settle for asking whether the interview successfully identifies the disorders as described in the manual. (p. 808)

Following this line of argument, even staunch critics of the DSM might agree that it presently serves as a nomenclature.

For those who are less critical of the DSM, there is little doubt that the recent versions of the DSM have been subjected to empirical scrutiny (Sanders, 2011). For example, the ordinary clinician who used the DSM-IV and DSM-IV-TR may not know that it underwent three stages of empirical testing to address the reliability, validity, and utility of its diagnostic categories (Frances, Mack, Ross, & First, 2000). Likewise, the research process that is being used to field test the DSM-5’s diagnostic categories has been made available to the public online in a paper entitled, “Protocol for DSM-5 Field Trials in Routine Clinical Practice Settings.” (APA, 2011).

That document presents a series of questions to be answered by its research concerning the DSM-5’s diagnostic criteria and two measures that are “incorporated into the diagnostic scheme for DSM-5,” namely, “diagnostic-specific severity measures” and “cross-cutting dimensional measures.” (p. 1). The questions that the DSM-5 field trials (APA, 2011) are designed to answer include precisely the questions that this dissertation will endeavor to answer; those relevant questions include:
**Diagnostic Criteria**

a. Are the proposed diagnostic criteria easy to understand and use by the clinician?

b. [Omitted.]

c. Are the proposed diagnostic criteria useful/helpful to the clinicians in planning treatments for patients? . . .

**Diagnostic-specific Severity Measures**

a. Are the proposed diagnostic-specific severity measures criteria easy to understand and incorporate into the clinical evaluation of patients?

b. Are the proposed diagnostic-specific severity measures reliable?

c. Are the proposed diagnostic-specific severity measures useful/helpful to clinicians in planning treatments for patients? (p. 1)

For the diagnostic criteria, question b., which was omitted above, concerned the validity of the proposed diagnoses, which is beyond the scope of this dissertation. For the diagnostic-specific severity measures, question b. concerns reliability. Both questions a. and c. concern the clinical utility of the proposed diagnoses. This dissertation’s research questions can be summarized by simply substituting “nomenclature” or “terms and definitions” (which are the two components of nomenclature; O’Carroll et al., 1996) for “diagnostic criteria” and “diagnostic-specific severity measures” in the DSM-5 field trials’ research objectives; also, the word “clients” is substituted for “patients” in recognition of the difference between the everyday language of counseling psychology and psychiatry. With those substitutions, the research questions to be addressed by this study are:

a. Are the proposed terms and definitions easy to understand and use by clinicians?

b. Are the proposed terms and definitions reliable?

c. Are the proposed terms and definitions useful/helpful to clinicians in planning treatments for clients?

With this formulation of research questions, b. concerns the reliability of a suicide-related nomenclature, while a. and c. concern the clinical utility of a suicide-related nomenclature. Since I will be comparing two major current nomenclatures, I also have a fourth research question, namely, “How do the two nomenclatures compare to each other with respect to their reliability and clinical utility?” This fourth research question will be answered through sound research design.
The Meaning of Reliability for Suicide-Related Nomenclatures

It would be helpful to describe what is meant by reliability in the context of suicide-related nomenclatures. In his book on measuring suicidal behavior, Goldston (2003) provided definitions of types of reliability as applied to suicide-related assessment: “[The] term reliability is used to refer to the ‘reproducibility’ of responses to a question or scale. Instruments are described and evaluated with respect to... interrater reliability...” (p. 17). Silverman et al. (2007a & 2007b) referred to the importance of establishing reliability, but left for future research actually collecting data regarding the reliability of that nomenclature. Posner et al. (2007) measured interrater reliability of nine suicidologists with each other, as well as with the judgments of pharmaceutical company raters. Comparing the existing work on these two suicide-related nomenclatures, then, only one of them intentionally collected data regarding reliability: the Columbia Classification Algorithm of Suicide Assessment (C-CASA) of Posner et al. (2007).

In terms of “reproducibility” described above, “reliability refers to the extent to which an experiment, test, or any measuring procedure yields the same results on repeated trials...” (Aboraya, Rankin, France, El-Missiry, & John, 2006). Interrater reliability refers to the reproducibility or stability of responses across raters, which is especially appropriate for comparing whether raters can classify vignettes into the same categories (as with DSM diagnoses) (Gutkind, Ventura, Barr, Shaner, Green, & Mintz, 2001). For interrater reliability, the question for this dissertation research would be whether different people classify identical vignettes into the same categories for the same nomenclature. For example, for the C-CASA nomenclature, the question would be whether different people could apply the C-CASA to the same vignette and come up with similar results.

This dissertation’s examination of reliability will be vignette based, since 26 vignettes can encompass all of the circumstances covered by the terms contained in the two competing nomenclatures. Some vignettes will describe self-injurious (but not suicidal) material, since both nomenclatures include competing terms for such material. In fact, this is a good, concrete example of the differences between the nomenclatures: each nomenclature uses different terms to describe a situation in which someone
hurts himself without suicidal intent. Silverman et al. (2007) would call that “self-harm,” while the C-CASA would call that “self-injurious behavior.”

For clarity and concreteness regarding how reliability will be assessed, it would be helpful to briefly describe the method that will be used to collect reliability data. First, a variety of clinicians (e.g., MFTs, psychologists, social workers, psychiatrists) would read a brief description of one of the two nomenclatures. Each clinician-participant will then receive 10 of the 26 vignettes (randomly selected) and will be asked which term from the nomenclature properly applies to the situation. Inter-rater reliability (i.e., kappa) will measure participants’ agreement with each other and with experts’ judgments concerning which term from the nomenclature properly described the vignette at issue.

**Shared and Unique Aspects of the Two Nomenclatures**

The research design of this dissertation will be more fully described in the third chapter (the method). However, at this early point it is important to note that this dissertation’s research design will focus on using short written vignettes; participants will decide, for each vignette, which term and definition of a particular nomenclature best applies to that vignette. Because these vignettes are so central to this dissertation’s research design, they are presented in Appendix B; this permits the reader to examine the vignettes whenever he or she would like. Appendix B also formally describes the method by which vignettes were created.

When the reader examines Appendix B and the vignettes contained therein, he or she will notice that each vignette is described in terms of shared aspects and unique aspects of each of the two nomenclatures. These shared and unique aspects were created by a content analysis of the overlapping and unique aspects of the two nomenclatures. Specifically, terms and definitions in each of the two nomenclatures are described in terms of the content that is unique to a particular nomenclature, as well as in terms of the content that is shared by the other nomenclature. These shared and unique aspects are summarized as letters A. through J. below, thereby representing nine categories of either shared or
unique aspects of the nomenclatures. These types of shared and unique aspects of the nomenclatures are listed immediately below:

A. Whether suicidal intent is present, absent, or unclear (shared by both nomenclatures).

B. Whether there are only thoughts present (e.g., suicidal ideation) (shared by both nomenclatures).

C. When there are only thoughts present, whether there are communications of those suicide-related thoughts to others (unique to the Silverman nomenclature, representing three terms within it).

D. Whether there are suicide-related behaviors present (shared by both nomenclatures).

E. Whether those behaviors result in injuries, no injuries, or fatal outcome (shared by both nomenclatures).

F. Whether the behavior is an accident (unique to C-CASA, representing one term within it).

G. Whether the behavior is only preparatory (unique to C-CASA, representing one term within it).

H. Whether there is a plan (unique to the Silverman nomenclature, representing three terms within it).

I. Whether there are suicidal threats (unique to the Silverman nomenclature, representing three terms within it).

J. Whether there is not enough information to determine if there was suicide-related behavior (unique to C-CASA, representing one term within it).

The purpose of constructing these categories A. through J. was to allow the construction of vignettes in an exhaustive but efficient manner. In other words, the 26 vignettes exhaustively represent every category in both nomenclatures; they do so by being constructed to match each of the categories A. through J. above. However, the vignettes were efficiently constructed because the unique categories were combined with shared categories within a single vignette.

For example, the concepts of suicidal communications, suicide threats, and suicide plan (which were each unique to the Silverman nomenclature) each represented three terms within the Silverman
nomenclature. The remainder of the terms and definitions contained within the Silverman nomenclature shared important conceptual aspects with at least one other term in the C-CASA. For example, both nomenclatures contain some version of the term “suicidal ideation,” the key aspect of which is that there are suicidal thoughts, as opposed to suicidal behaviors. Likewise, both nomenclatures contained the concept of suicidal intent, as well as whether the suicidal intent was present, absent, or unclear whether it existed in a particular instance.

Therefore, when the reader examines the 26 vignettes reviewed in Appendix B, he or she will see that prior to each vignette, the italicized section details which categories (A. through J.) are present, absent, or unclear (in the case of suicidal intent). This permits the careful reader to verify that the 26 vignettes in fact exhaustively and efficiently represent all the terms within both nomenclatures. However, to be able to verify this thoroughness requires a detailed understanding of the terms and definitions of the two nomenclatures (presented in Appendix A). A formal description of how vignettes were constructed is available in Appendix B, prior to the presentation of the vignettes themselves.

The Distinction between Validity and Reliability for Suicide-Related Nomenclatures

The measurement of interrater reliability between clinician-participants and expert raters – in which expert raters are suicidologists and/or co-authors of the two competing suicide-related nomenclatures – could be thought of as pertaining to validity rather than reliability. Goldston (2003) described the difference between reliability and validity, and characterized validity as referring to accuracy: “It is important to note that questions or scales may be ‘reliable’ or yield reproducible responses without necessarily being ‘valid’ or accurate measures of what it is they really purport to assess.” (p. 17). As with reliability, the “instrument or scale” at issue would be a written summary of a particular suicide-related nomenclature (e.g., the Silverman nomenclature) that is given to participants to apply to hypothetical vignettes. The accuracy of what the Silverman nomenclature purports to assess would constitute its validity.
As applied to suicide-related nomenclatures, validity therefore refers to the accuracy of the existence of what the nomenclature purports to describe: namely, real-world thoughts and behaviors corresponding to the terms and definitions laid out in the nomenclature. Mullins-Sweatt and Widiger (2009) put the point more generally in their description of construct validity: “Construct validity is concerned with determining the truthfulness, reality, or existence of a construct.” (p. 302). The constructs at issue in suicide-related nomenclatures are terms and definitions; the truthfulness, reality, or existence of the construct described by each term (and its corresponding definition) relates to whether the concepts described and defined in each of the terms actually exist in the real world of people experiencing suicidal thoughts and behaviors.

With this more thorough understanding of validity as it applies to suicide-related nomenclatures, let us return to the practical question of whether kappas between expert raters and clinician-participants should be considered to relate to validity or reliability. The above description of validity as referring to the existence, in the world, of the referents of the definitions and terms contained in a nomenclature, makes clear that the agreement or disagreement between expert raters and clinician-participants relates to reliability, not validity. Specifically, suicidologists or co-authors of the nomenclatures would be experts in how to apply the nomenclature to a vignette, but the clinician-participants would actually be more expert regarding whether the referents of the nomenclature’s terms exist in the world (since that would refer to clinical data, which is more their purview).

In short, if this dissertation were to explore the validity of the two major suicide-related nomenclatures, we would rather ask a different research question than the ones that I am in fact pursuing: namely, whether clinicians find that the terms and definitions accurately reflect or capture their clients’ symptom presentations. To be fair, this would not capture the entire range of potentially valid suicide-related terms and definitions: for example, a particular suicide-related nomenclature might specify terminology and definitions which match suicide-related thoughts or behaviors existing in the world, but no one in the sample of participants has had a client whose symptoms match the term and
definition. We could try to expand the pool of real-world examples by specifying that clinicians are to remember whether they have ever had a client who matched the term at issue. We could try to expand the pool of examples further by asking whether clinician-participants expect that they would eventually encounter a client with symptoms that fit the definition of the particular term under consideration. However, the next section demonstrates that to do so would actually be a question related to the clinical utility of the nomenclature, not its validity. The essential point to recognize, however, is that validity of suicide-related nomenclatures is a complex area that is best left to future research.

**The Distinction between the Validity and the Clinical Utility of a Nomenclature**

In contrast, the clinical utility of the two competing suicide-related nomenclatures is a critical research question for this dissertation. Research related to the development of DSM diagnostic categories is again helpful, this time in elucidating the difference between validity and clinical utility of a nomenclature. Kendell and Jablensky’s (2003) article in the *American Journal of Psychiatry* provided an incisive overview of the difference between validity and clinical utility. Like the authors cited in the preceding section on validity, they define diagnostic validity as having to do with the existence of distinct categories in the world; they convincingly argue that the validity of a diagnostic category should either “be defined by a physiological, anatomical, histological, chromosomal, or molecular abnormality,” or else, “if the defining characteristic of the category is a syndrome, this syndrome must be demonstrated to be an entity, separated from neighboring syndromes and normality by a zone of rarity.” (Kendell & Jablensky, 2003, p. 8). While I will address their argument in more detail in my literature review (Chapter 2), it can be briefly summarized as stemming “from several reasons why the crucial issue in determining validity is not understanding of etiology but rather the existence of clear boundaries or qualitative differences at the level of the defining characteristic” of a diagnostic category.” (p. 8). In fact, they provide four such reasons, the most important of which for our purposes is their third reason:

Third, most psychiatric disorders... are still defined by their clinical syndromes because their etiology is still largely unknown. It would be perverse to define validity such that no syndrome of unknown etiology could be accepted as a valid category, even if discriminant function analysis demonstrated that such a syndrome is separated by a
zone of rarity from neighboring syndromes. Such evidence of a natural boundary would strongly suggest – but not prove – that the etiology of that syndrome was different from that of its neighbors. (p. 8)

Just as the etiology of most psychiatric disorders is currently unknown, the origins of suicidal thoughts and behavior are similarly opaque given the current state of the science of suicidology. Far from seeing this reality as indicative of a hopeless state of affairs, Kendell and Jablensky point out that psychiatry and suicidology are currently in the position “that most of medicine was in 200 years ago.” (p. 9). Rather than bemoan the fact that we are a long way from establishing the validity of our science’s categories, they suggest that we recognize that it is not validity but rather clinical utility that should currently guide the adoption or abandonment of particular categories.

It is a truism that reliability is necessary for validity (Robins & Cottler, 2004), and reliability is relevant to clinical utility, though not totally necessary for it in all cases (Kendell & Jablensky, 2003).

Kendell and Jablensky (2003) explain:

Validity... is an invariante characteristic... in principle a category cannot be partly valid. Either it is or is not valid, and its validity does not depend on the context. Utility, on the other hand, is a graded characteristic that is partly context specific... Bipolar disorder may be a very useful concept in an acute admission unit, where it is important to distinguish between psychotic states that are and are not induced by stimulant drugs... However, it may be less useful in a rehabilitation program where the crucial issues may be which residents take their medication regularly, which are likely to have psychotic relapses, and which ones upset the others. (p. 10)

The upshot of this distinction is that whether reliability is key to clinical utility depends on the context in which the clinical utility is explored. In the context of suicide research, in which different definitions and terms have seriously hampered research and clinical practice, I contend that reliability is critical to clinical utility. However, reliability is not necessary for utility in other situations, unlike with validity.

Likewise, for judgments about which terms and definitions to adopt in a suicide-related nomenclature, I contend that issues of clinical utility are currently more relevant to improving the state of our nomenclature than are those of validity. My argument rests on three central components of clinical utility described by Mullins-Sweatt & Widiger (2009): “ease of usage, communication, and treatment planning.” (p. 303). I hypothesize that the ease of usage of the terms of a nomenclature will bear a direct
relationship to whether those terms are or are not adopted for widespread use by busy clinicians. In this, for example, I agree with the Silverman nomenclature’s decision to replace the O’Carroll nomenclature’s term, “instrumental suicide-related behavior,” with the easier-to-use term, “suicide-related communication.” However, I go further than the Silverman nomenclature in that I guess that we will never accomplish the widespread adoption of complex terms such as “suicide-related” anything, whether it is “suicide-related behavior,” “suicide-related communication,” or the like. I believe this despite the fact that I use “suicide-related nomenclature” to describe the two current major nomenclatures. I use “suicide-related” rather than the easier-to-use “suicide nomenclature” throughout this dissertation because the latter seems less accurate than the former: prominent researchers make strong arguments based on empirical data that suicide-related nomenclatures ought to capture self-injurious behavior in which the person had no suicidal intent (Nock et al., 2007). They base their argument on the fact that there is an empirical correlation between self-harming behaviors that lack suicidal intent (such as cutting, self-hitting, or burning) and suicidal behaviors like attempted and completed suicide (Nock et al., 2007). Because self-injurious but non-suicidal terminology should thus be included in a suicide nomenclature, it seems more accurate to describe such nomenclatures as “suicide-related nomenclatures.” However, what is most important now for the state of suicide research and clinical practice is to identify a set of terms that most clinicians can agree to use; therefore, my opinion and the opinion of other suicide experts is less critical than the empirical evidence on clinician-participants’ utility preferences, which I will gather for this dissertation.

The second component of clinical utility as elaborated by Mullins-Sweatt & Widiger (2009) is communication. I have already argued that a standard nomenclature will provide a common language for clinicians and researchers in suicidology, thereby promoting meaningful comparison between research studies and improved understanding between clinicians consulting with each other about cases. I now assert that the development of some standard nomenclature – that is, any nomenclature that can garner widespread acceptance by both clinicians and researchers – is more critical at this point in history than
whether that nomenclature is the “correct” one, which is properly understood as an issue of validity. I base this claim upon the aforementioned state of suicidology: our science is not currently in a position to know which terms and definitions would be valid with respect to their matching categories that exist in the world. Suicidology is still too young a science for valid categories to be within reach. However, we desperately need some standardization of terminology and definitions so that clinicians and researchers are at least talking about and studying the same things.

The third component of clinical utility of a suicide-related nomenclature is whether the terms and definitions contribute to treatment planning. Here I echo the final line of Mullins-Sweatt’s and Widiger’s 2009 article: the purpose of improving suicide-related nomenclature “is not simply a matter of addressing a scientific interest in understanding and explaining” suicide phenomena; rather, “its ultimate purpose is to help reduce pain and suffering within the general population – more specifically, to facilitate the practice of clinicians administering clinical care.” (p. 310). As such, it makes sense to prefer to select terms and definitions that clinicians think will contribute best to effective treatment planning with suicidal individuals; all other things being equal, we thereby increase the likelihood that we will help reduce the considerable pain of clients suffering from suicidal thoughts and behaviors.

To assign greater importance to the collection of data related to the clinical utility of suicide-related nomenclatures is not to conceptually elevate clinical utility above validity; in contrast, as Mullins-Sweatt and Widiger stated, “clinical utility and construct validity are not entirely independent concepts. An entirely invalid or illusory diagnostic construct would have no clinical utility, and an entirely valid diagnostic construct would have at least some clinical utility” even if that utility were simply the fact that we could accurately describe the hopeless clinical condition to clients, their families, and policymakers (2009, p. 310). However, as Samuel and Widiger (2011) emphasized, “even the most valid model would fail in its purpose if not used effectively within clinical practice.” (p. 14).
Summary of the Research Questions

The purpose of this dissertation research is to provide data regarding the reliability and utility of the two major current suicide-related nomenclatures: the Silverman nomenclature and the C-CASA nomenclature. Clinicians’ use of these nomenclatures depends on clinicians’ accurate understanding of the meaning of their terms, which is an internal process that is not easily observed in a research project. Partly, we can gauge clinicians’ accurate understanding of the nomenclature through the outcomes of their application of it to particular vignettes.

Therefore, it would be valuable to ask clinician-participants directly about their experience of using the nomenclature itself. What did they like most about the nomenclature? What did they like least? Collecting data about clinicians’ experience of using the nomenclatures can help us to evaluate what about a particular nomenclature is promising or disappointing. Analysis of themes contained in free-form written responses to those types of questions can also give us hints about promising future directions for terms and definitions. This data would be very different than the numerical Likert-type responses collected for analysis of reliability kappas and clinical utility questionnaires.

Questions of clinicians’ perception of the utility of the nomenclatures are central to this research. Part of what we want in a suicide-related nomenclature is that clinicians will be willing to use it. If terms and definitions are going to contribute to a common language for mental health professionals, then clinicians must think that the benefits of using the nomenclature outweigh the costs of learning and applying it. I would also like to collect future-oriented feedback from clinicians regarding the nomenclatures, including whether they believe the nomenclature would help them in their day-to-day practice.

While I hope that the specific content of the results of data analysis detailed above will add incrementally to the improvement of suicide-related nomenclature, my biggest research goal does not relate to the content of the results. It is most exciting to me that this dissertation might offer an example of collecting and reporting clinician-level data regarding the reliability and utility of particular suicide-
related nomenclatures. If this dissertation is successful, I hope that this methodology can be replicated by other researchers, thereby contributing to a truly scientific study of suicide-related nomenclature. I seek to shift the paradigm of data collection and reporting from one that is based on an expert’s summary of data that cannot be seen by the journal reader (typically collected from workgroups or informal clinician feedback), to one that is based on the collection of traditional psychological-scientific data. This data would be available to others in published journal articles in which the research procedure is detailed sufficiently that it could be replicated by independent researchers. In that way, I hope that this dissertation contributes to the advancement of psychological science in the study of suicide-related nomenclatures.
CHAPTER II
LITERATURE REVIEW

Introduction

The first chapter briefly summarized the state of research with respect to the two competing suicide-related nomenclatures. This chapter will expand upon that to comprehensively review the extant literature for each of the nomenclatures. Furthermore, this chapter will examine the types of evidence typically cited in support of the reliability of suicide-related nomenclatures – namely, that collected through expert workgroups – versus that of the evidence cited in support of the reliability of medical- and DSM-related nomenclatures, which is typically the interrater reliability of the nomenclature. The DSM, as the predominant mental health nomenclature, provides a blueprint for both collecting interrater reliability evidence and for collecting evidence of the clinical utility of a nomenclature.

Because it has more citations to the entirety of its nomenclature, research relevant to the Columbia Classification Algorithm for Suicide Assessment (C-CASA; Posner et al., 2007) will be presented first. That review will be followed by a review of Silverman et al. (2007a & 2007b), which has had a primarily academic impact. The practical effects of the C-CASA will be highlighted while the theoretical appeal of the Silverman nomenclature will be focused upon.

Literature Review of the Columbia Classification Algorithm of Suicide Assessment (C-CASA)

There is a substantial literature concerning the Columbia Classification Algorithm of Suicide Assessment (C-CASA). The history of the C-CASA will be presented first, followed by an overview of the C-CASA system. Next, articles that used the whole C-CASA system itself will be reviewed. Articles will then be reviewed in which C-CASA was cited for only one or a few definitions, followed by those articles which simply praised C-CASA as a system. Finally, the few articles that criticized C-CASA will be addressed.
History of the Food and Drug Administration and C-CASA

In 1997, the United States Food and Drug Administration ("FDA") Modernization Act instituted monetary rewards for pharmaceutical companies who set up pediatric randomized controlled trials ("RCTs"); this created a wealth of data regarding selective serotonin reuptake inhibitors ("SSRIs") and their effects on children and adolescents (Wu, 2009). As controversy escalated concerning the safety of SSRIs and their possible increase of suicidal thoughts and behaviors, the FDA became aware that drug companies may have miscoded suicidal thoughts and behaviors (Wu, 2009). For example, records indicating that a teenage girl “slapped herself in the face” was erroneously labeled a suicide attempt, as was another instance in which a teenage boy reported suicidal ideation but lacked any intention to act on his thoughts (Gangwisch & Jacobson, 2009, p. 372). This led the FDA, in 2003, to commission a study in which neutral investigators at Columbia University came up with a system for re-rating pediatric suicide-related data, otherwise known as adverse events (Gangwisch & Jacobson, 2009). The Columbia Classification Algorithm of Suicide Assessment ("C-CASA") was the system devised as a result of the FDA’s request.

Manalai, Woo, and Postolache (2009) described the state of the FDA and its approach towards C-CASA:

The [FDA] division evaluating the psychiatric safety of medications has required inclusion of the Columbia Classification Algorithm for Suicide Assessment (C-CASA) in retrospective studies and Columbia-Suicide Severity Rating Scale in prospective studies of any drug with possible neuropsychiatric side effects. However, other divisions of the FDA still do not have a procedure in place for using a more standardized scale for assessing suicidality. (p. 277)

The Columbia-Suicide Severity Rating Scale ("C-SSRS") is a prospective assessment instrument that “maps to the C-CASA coding system...” (Gassmann-Mayer et al., 2011, p. 555). In other words, the FDA requires drug companies to use C-CASA when reporting adverse events if their medication potentially has neuropsychiatric side effects. However, for other types of medication side effects, there is still no standardized classification or nomenclature for suicide-related
adverse effects. Many classes of drugs have neuropsychiatric side effects, though, making C-CASA an important and widespread nomenclature and classification system.

**Overview of the Columbia Classification Algorithm of Suicide Assessment (C-CASA)**

Posner et al. described their suicide-related nomenclature in their 2007 article entitled, “Columbia Classification Algorithm of Suicide Assessment: Classification of suicidal events in the FDA’s pediatric suicidal risk analysis of antidepressants.” The first important characteristic to note is that the C-CASA is primarily a classification system and only secondarily a nomenclature. While it defines a nomenclature explicitly, it was created in order that “a standardized suicidal rating system [could provide] data for the pediatric suicidal risk analysis of antidepressants conducted by the Food and Drug Administration (FDA).” (Posner et al., p. 1035). In other words, C-CASA was constructed for a practical rather than theoretical purpose: to help determine whether antidepressant medication increased children’s and adolescents’ suicide risk. However, this does not mean that creating terms and definitions (i.e., a nomenclature) was not one of its purposes, albeit not its primary one: one of the goals of the C-CASA was “to develop more specific definitions for suicidal behavior. In particular...to differentiate suicide attempts with at least some intent to die and self-injurious behavior with no intent.” (Emslie, Ventura, Korotzer, & Tourkodimitris, 2009, p. 722).

Because practical assessment of dangerousness was its purpose, the C-CASA was centered around so-called “adverse events” from pharmaceutical clinical trials; Posner et al. (2007) reviewed 427 adverse events from 25 pediatric clinical trials that had been “systematically identified” by drug companies for the FDA’s safety review (p. 1035). Utilizing C-CASA, the FDA identified 26 new suicidal events (defined as either suicidal behavior or suicidal ideation) that pharmaceutical companies had not previously identified as potentially suicide-related. Additionally, C-CASA resulted in 12 adverse events being reclassified as not suicidal, for a total of 38 differences between drug companies’ and the FDA’s analyses.
C-CASA’s reliability was judged utilizing interrater reliability; nine expert suicidologists were used as raters. Experts were “internationally recognized experts in suicide and suicide assessment” (Posner et al., 2007, p. 1038) who had ties neither to Columbia University (whose investigators designed the study) nor to the FDA (which implemented the procedures designed by Columbia investigators). Posner et al. (2007) reported excellent overall reliability of median C-CASA concordance between the nine experts: the intraclass correlation coefficient (“ICC”) was .89 among the expert raters, across all seven categories of the C-CASA represented (there were no completed suicides).

However, there was variability among the seven categories represented within the C-CASA: ICCs ranged from .47 (for the category “Not Enough Information”) to .97 (for the category, “suicidal ideation”). Other problematic categories included “self-injurious behavior, no suicidal intent” (ICC = .59) and “self-injurious behavior, suicidal intent unknown” (ICC = .67); in contrast, categories with good-to-excellent reliability included “Other, no deliberate self-harm,” (ICC = .93); “preparatory acts toward imminent suicidal behavior” (ICC = .89) and “suicide attempt” (ICC = .81) (Posner et al., 2007, p. 1040). This between-category variability highlights the importance of isolating potentially problematic terms and definitions that might have contributed to the modest reliabilities of some categories. Overall, though, suicide-related categories displayed good agreement between the expert suicidologist raters and drug company raters (kappa = .77).

This overall positive result mirrored the generally positive stability of the C-CASA itself. The FDA’s own audit of the C-CASA found a kappa of .84 and percent agreement of 89 percent between the FDA’s raters and Columbia’s raters; this was despite an intentional over-sampling of 31 difficult-to-rate cases (Iyasu, 2004). Those 31 tough cases were selected from among 64 of the 423 adverse event cases reviewed by Columbia and reported in Posner et al. (2007); the hard cases represented all events “reclassified by Columbia to non-suicidal or other events (n=2)”; as well as one third of cases (i.e., 10) “newly identified and classified by Columbia as possibly suicide related (n=29);” and one third of cases (i.e., 19) “difficult to classify [which was] defined as events with discordant initial independent ratings by
Columba reviewers (n=56).” (Iyasa, 2004, slide 6). Finally, the FDA audit sampled ten percent of cases (i.e., 33) that had high concordant initial ratings by Columbia reviewers (n=336).

Posner et al. (2007) has been an influential article; it has been cited 141 times since its publication only five years ago, according to the Social Science Citation Index (Thomas Reuters, 2012). Furthermore, Gassmann-Mayer et al. (2011) asserted that “the FDA has endorsed the Columbia Classification Algorithm of Suicide Assessment (C-CASA) and encourages the use of a scale that maps to the C-CASA coding system...” (p. 555). Perhaps because of the FDA’s endorsement, citations of Posner et al. (2007) and the C-CASA have proliferated in recent years.

**Articles Using C-CASA Itself**

Vanderburg, Batzar, Fogel, and Kremer (2009) reported on 126 studies containing almost twenty thousand (N=19,923) patients who were treated with sertraline or placebo. Using C-CASA to classify suicide-related adverse events, ninety-nine suicidal events were identified, including four completed suicides. Those suicidal events were classified into nine categories. C-CASA itself contains only eight categories, but the FDA split C-CASA’s category “Other, no deliberate self-harm” into two categories: “not enough information (fatal) (code 6)” and “not enough information (nonfatal) (code 9)” (Vanderburg et al., 2009, p. 676). Otherwise, the FDA’s nine codes were identical to C-CASA’s eight categories: “completed suicide (code 1); suicide attempt (code 2); preparatory acts toward imminent suicidal behavior (code 3); suicidal ideation (code 4); self-injurious behavior, intent unknown (code 5). . . self-injurious behavior, no suicidal intent (code 7); other – accident, psychiatric, medical (code 8). . .” (Vanderburg et al., p. 676).

These nine codes were used throughout retrospective FDA-required research into suicidal adverse events. Thus, C-CASA has been utilized in literally hundreds of clinical trials in the United States, as a method of reporting suicide-related adverse events.

Hesdorffer and Kanner (2009) reported on C-CASA’s use in the context of anti-epileptic drugs: “The C-CASA suicidality classification was used in 199 trials for the treatment of epilepsy (31%), pain (41%), and psychiatric disorders (28%).” (p. 980). Hesdorffer and Kanner (2009) also broke down the
results into the four types of suicidal behavior contained in the C-CASA system: “ideation was the most frequent... and accounted for 67.6% of all suicidality. Preparatory acts accounted for 2.8%, attempt for 26.8%, and completed suicide for 2.8%.” (p. 980). Therefore, almost 95 percent of suicidal events were either suicidal ideation or suicide attempt, with less than three percent each for completed suicide and preparatory acts toward imminent suicidal behavior.

Blasczyk-Schiep, Kazén, Kuhl, and Grygiflski (2011) studied whether the Rorschach test could differentiate higher suicide risk from lower suicide risk among a sample of adolescents and young adults. In so doing, they used the C-CASA itself, specifically for three of the four categories related to suicidal events (there were no completed suicides in their sample, so that category was not relevant). Blasczyk-Schiep et al. wrote, “The inclusion criterion for the study was the C-CASA classification category for suicidal events with suicidal intent...” (2011, p. 520); they looked at 116 inpatients who met those C-CASA criteria. Rorschach scores identified 69% of those with suicide attempts.

Tourian, Padmanabhan, Groark, and Ninan (2010) assessed whether the antidepressant desvenlafaxine was correlated with higher risk of suicidal thoughts and behavior. In so doing, the authors utilized the Columbia Classification Algorithm of Suicide Assessment (C-CASA) to rate nine different studies of desvenlafaxine. In those nine studies, there were 28 adverse events reported, 15 of which were either suicide attempts, completed suicide, or suicidal ideation (preparatory acts were not present). The authors did not find evidence of increased risk of suicidality.

Kennard et al. (2009) studied whether use of placebo for 12 weeks followed by 24 weeks of active treatment was better, worse, or the same as 36 weeks of active treatment. Kennard et al. used C-CASA for classifying and defining suicidal events: “Primary suicidal events were defined using the Columbia Classification Algorithm of Suicidal Assessment and coded independent of knowledge of treatment or course by the Columbia University Suicidlity Classification group.” (2009, p. 338). No difference was found between suicidal events in initial placebo versus totally active treatment groups.
Ougrin et al. (2012) used the C-CASA to look at whether adolescents with suicidal self-harm differed compared to adolescents with non-suicidal self-harm. Specifically, Ougrin et al. classified an index self-harming episode using the C-CASA to differentiate that episode into either suicidal (i.e., defined as a suicide attempt) or non-suicidal (defined as self-injurious behavior, either with no suicidal intent or with unknown suicidal intent). The two senior psychiatrists who classified index episodes into suicidal versus non-suicidal categories displayed “a good level of agreement… (κ = .69, p < .001),” (Ougrin et al., p. 15) providing further evidence of the reliability of C-CASA.

Crocq et al. (2010) used the C-CASA to assess adverse events – specifically suicide attempts – prospectively in a sample of schizophrenic patients who were treated with risperidone or sertindole. The FDA requested post hoc analysis using the C-CASA as part of a larger study (Peuskens et al., 2008); adverse attempts that Crocq et al. (2010) had classified as suicide attempts were reviewed and were classified using one of three C-CASA categories: completed suicide, suicide attempt, or preparatory acts toward imminent suicidal behavior.

Another study looking at antipsychotic medication was Karayal, Anway, Batzar, and Vanderburg (2011). They addressed “suicidality” in RCTs of ziprasidone (Geodon®); suicidality was defined as encompassing “suicidal ideation,” “preparatory acts toward imminent suicidal behavior,” “suicide attempts,” and “completed suicide” (p. 370). The C-CASA was the main outcome measure in Karayal et al.’s study. Regarding the C-CASA, the authors concluded that “the C-CASA classification system provides a currently acceptable method for retrospective identification of treatment-emergent suicidality and suicidal behavior.” (Karayal et al., p. 373). In other words, they used the C-CASA to categorize adverse suicide-related events and did not find it lacking – they recommended it for future use.

Wadden et al. (2010) utilized the C-CASA in a randomized trial of lifestyle modification, a cannabinoid receptor inverse agonist, and placebo for maintaining weight loss. In their section regarding safety assessment, the authors wrote:

Adjudication of possibly suicide-related adverse events was conducted according to U.S. Food and Drug Administration guidance for exploring suicidality (i.e., suicidal ideation
and behavior) in placebo-controlled clinical trials. Narratives for these adverse events were written, and adjudication was performed by an independent, external committee that used the Columbia Classification Algorithm of Suicide Assessment. (p. 2303)

In other words, utilization of C-CASA seemed to be perceived as the FDA-approved standard by which suicide-related adverse events were to be judged. As such, Wadden et al. (2010) applied C-CASA itself in a large-scale randomized clinical trial that had nothing, on its face, to do with suicide. Rather, the logic of Wadden et al. suggested that anytime there might be an adverse event involving suicide, the FDA would look kindly on a study in which C-CASA was used. Similar studies regarding obesity and the same cannabinoid receptor inverse agonist (Proietto et al., 2010; Kipnes et al., 2010; Aronne et al., 2010) used identical language and had overlapping authorship to that of Wadden et al. (2010), suggesting that those authors likewise used C-CASA according to the same logic. A study involving obesity and a cannabinoid receptor antagonist made this relationship between C-CASA and the FDA quite clear. Van Gaal, Pi-Sunyer, Després, McCarthy, and Scheen (2008) wrote, “The Columbia Classification Algorithm of Suicide Assessment is recommended by the Food and Drug Administration for the assessment of suicide risk and comprises nine categories...” (p. S237). Like the other studies mentioned above concerning obesity and cannabinoid receptors, Van Gaal et al. used the C-CASA to classify its suicide-related adverse events.

A multi-site randomized clinical trial in Japan of newer antidepressants for depression stated that it used the C-CASA “to allow more detailed assessment of suicidality” than was provided by the adverse event manual of the Japanese Ministry of Health, Labour and Welfare (Furukawa et al., 2011, p. 129). Similarly, Woldu et al. (2011) studied adverse events related to adolescent antidepressant use or non-use through utilization of “the Brief Scale for Rating Severity of Suicidal Behavior (B-SSRS) with categories classified according to the Columbia Classification Algorithm for Suicide Assessment.” (p. 492). In a related study, Brent, Melhem, et al. (2010) likewise used the C-CASA via the B-SSRS to study depression, also in the same sample of adolescents as that of Woldu et al. (2010). Finally, Vitiello et al. (2009) also used C-CASA to study suicide-related adverse events in research called the Treatment for Adolescents with Depression Study (TADS), which examined fluoxetine, cognitive-behavioral therapy, and their combination
over 36 weeks. About ten percent of the sample (44 of 439 adolescents) either had suicidal ideation or made a suicide attempt, but no one died by suicide. The study found no evidence that taking antidepressant medication led to an increase in suicidal events.

Morris and Trivedi (2011) stated that a measure they created, the “Concise Health Risk Tracking” (CHRT) instrument, “allows clinicians to code suicidality using the Columbia Classification Algorithm of Suicide Assessment (C-CASA) criteria.” (p. 448). However, examination of their CHRT measure revealed that while it followed the general categories of the C-CASA, terms and definitions were either not explicit or varied from the C-CASA itself. For example, suicide was not defined and was referred to as “Completed Suicide” (a term which C-CASA does not use). On the other hand, suicide attempt was defined as “a potentially self-injurious behavior associated with intent to die. Intent can be stated by patient or inferred by rater.” (Morris & Trivedi, 2011 p. 454); this is a paraphrase of the C-CASA’s definition which, while generally accurate, does not contain the phrase “some intent to die,” (emphasis added) which is an important component of C-CASA’s definition of suicide attempt (Posner et al., 2007, p. 1037).

C-CASA has been used to classify suicide-related adverse events for allergy medication as well. Philip et al. (2009) studied whether montelukast, an allergy medication, caused suicidality as defined by C-CASA. Just as with antidepressants and antipsychotic medication, “adjudication of PSRAEs [Possibly Suicidality-Related Adverse Events] with the Columbia Classification Algorithm of Suicide Assessment (C-CASA) was performed by an independent Suicide Classification Center (SCC, Columbia University, New York, NY) blinded to treatment assignment.” (Philip et al., p. 692). The research found no significant suicidality in either placebo groups or clinical groups, across 41 studies put forth by the pharmaceutical company Merck. There was only one patient with suicidal ideation in the montelukast group. Manalai, Woo and Postolache (2009) reported similar results for montelukast and an anti-asthma medication called zafirlukast, suggesting that neither that allergy medication nor that asthma medication causes suicide-related thoughts or behavior.
In summary, a large number (N = 20) of studies have used the C-CASA in its entirety. These studies have ranged across psychiatric, psychological, obesity-related, and substance abuse topics. In general, studies have found the C-CASA to be a helpful and useful tool to operationalize suicide-related nomenclature and classification.

**C-CASA Used for Only One or a Few Definitions**

A number of studies used the C-CASA to create one or several definitions that were consistent with C-CASA, though they did not use the entirety of the C-CASA. For example, in a study that looked at whether suicidal intent and lethality predicted subsequent suicide attempts, “consistent with contemporary systems for the classification of suicidal behaviors [Posner et al., 2007; Silverman et al., 2007], a suicide attempt was defined as a potentially self-injurious behavior associated with at least some degree of psychological intent to end one’s life.” (Sapyta et al., 2012, p. 224).

Seemüller et al. (2008) used C-CASA to analyze the link between antidepressants and suicidal thoughts and behavior in adults. However, they used C-CASA only to classify suicides and suicide attempts; they chose to use another instrument to monitor suicidal ideation which allowed for recording changes over time. This study was noteworthy because it used C-CASA prospectively, rather than as it was designed (retrospectively).

Likewise, Spoletini et al. (2011) used C-CASA for only one definition, which was also a suicide attempt. Though C-CASA was cited for that definition, the definition itself was not the one contained in C-CASA; instead, suicide attempt was defined as, “Self-injurious behavior associated with the intent of dying.” (Spoletini et al., p. 207). Pan et al. (2011), in their study of depressed adolescents, stated that they used C-CASA’s definition of suicide attempt to separate suicide attempters from non-attempters. A study of adverse events among depressed adolescents (Brent, Emslie, et al., 2009) used C-CASA’s definition of suicide attempt and non-suicidal self-injury but did not explicitly define suicidal ideation and included the term (not used by C-CASA) “suicidal threat.” (p. 420). In other words, they followed part of C-CASA’s terminology but not all of it.
In a similar vein, Contreras et al. (2010) defined “suicide behavior [sic] according to The Columbia Classification Algorithm of Suicidal Assessment requirements (self-directed act documented in any of the available sources of information)” (p. 85). No more information than the above was provided since the study itself was primarily about genotypes (i.e., a biologically-focused investigation), so it is difficult to know how closely the authors stuck to the C-CASA’s definition of suicidal behavior. However, the C-CASA does not use the term “suicide behavior” so that fact by itself is not encouraging.

Cutler et al. (2011) was an example of a study that stated that it used a system similar to the C-CASA, though not the C-CASA itself: “The relative risk of suicide for both treatment groups was estimated using a suicidality classification system similar to the one established by Columbia University [Posner et al., 2007]. Suicidal behavior/ideation was defined as a Columbia classification score of 1, 2, 3, or 4, whereas potential suicide events were included under a Columbia classification score of 5, 6, or 9.”

Articles that Praised C-CASA

A number of citations simply referred to C-CASA in positive terms. Mann et al. (2009) wrote a review article regarding suicide and genetics that provided a nice summary of the ascendancy of C-CASA in the psychiatric literature:

Comparison of findings between studies is limited by lack of uniform definitions of suicidal ideation and behaviors. The Columbia Classification Algorithm of Suicidal Assessment is being widely adopted and provides operational definitions of suicide and suicide attempt. . . (p. 556)

The apparent reason for its widespread adoption in psychiatric circles is that its classification system and terminology are highly regarded as progress in the field. For example, Walkup, Townsend, Crystal, and Olfson (2012) referred to C-CASA when they stated, “Progress has recently been made in the development of a scale to assess suicidal behavior as an adverse event in clinical trials [Posner et al., 2007],” (p. 180). As another example, Hammad, Pinheiro, and Neyarapally (2011) referred to Posner et al. (2007) in the course of an article about method difficulties with randomized controlled clinical trials; the purpose of citing C-CASA was to note that adverse events are so unpredictable that it led Posner et al. to create “an evaluation scheme to search and adjudicate suicidality events,” which resulted in finding a
significant number of misclassified adverse events (Hammad, Pinheiro, & Neyarapally, 2011, p. 562). As a result, “the Columbia Classification Algorithm of Suicide Assessment (C-CASA) was found to be useful and necessary.” (Valtonen et al., 2009).

Editorials also praised the C-CASA. For example, Larkin and Beautrais (2010), in their review of suicide-related problems in emergency departments, noted the challenges presented by the lack of agreed-upon nomenclature for suicidology. They wrote, “There is no agreed-upon terminology or operational definition of suicidal behavior in English-speaking countries, let alone the rest of the non-English-speaking world.... The classification algorithm for assessment of suicidal behavior, developed by Posner and colleagues, begins to address this issue by offering standardized definitions.” (Larkin & Beautrais, 2010, p. 4). Finally, Manalai, Woo and Postolache (2009) offered in their “expert opinion” review article that C-CASA was a “well-validated tool [to] aid in accurately defining and classifying suicidality while eliminating confounders.” (p. 279).

In their essay on secondary prevention of suicide, Ganz, Braquehais, and Sher (2010) recounted C-CASA as “recent retrospective research [that] delineates seven distinct categories of ‘suicidality’: (1) completed suicide, (2) suicide attempt, (3) preparatory acts toward imminent suicidal behavior, (4) suicidal ideation, (5) self-injurious behavior without intent to die, (6) nondeliberate self-harm, (7) self-harm behavior with unknown suicidal intent.” (p. 1). This accurate summary of C-CASA implicitly endorses the C-CASA system since there are no other categorization systems cited by Ganz et al. (2010). In contrast, Heilbron, Daniel, Compton, and Goldston (2010) cited several classification systems but referred positively to C-CASA throughout their article on why the term “suicide gesture” should no longer be part of the lexicon of suicidology.

Beyond positive reference to C-CASA, though, there were also articles that cited it as a model of methodology. In their review of adverse events in substance abuse treatment, Mansbach, Schoedel, Kittrelle, and Sellers (2010) praised the C-CASA and remarked that a similar process relying on subject matter experts would benefit the substance abuse treatment field, stating, “The establishment of a
classification system on the C-CASA model would be a service to the field and [would] improve the utility of [adverse event] analysis as a component of the [substance] abuse potential assessment.” (p. 175).

Mansbach et al. therefore suggested that analysis of adverse events in substance abuse treatment follow the C-CASA model of narrative review of blinded vignettes, by subject matter experts, in order to determine which types of treatments yielded the fewest and least serious adverse events.

In summary, the words of Kilbane, Gokbayrak, Galynker, Cohen, and Tross (2009) – in their study of panic, bipolar disorder, and suicide – seem particularly appropriate regarding the problems of definition and the way in which C-CASA attempts to make a contribution to the field:

One particular limitation is the variability and inconsistency present in the use of terms relating to suicidal ideation, behavior, attempt, and completion. This limitation is highlighted in the work of Posner et al. (2007), and future use of a standardized assessment tool such as the Columbia Classification Algorithm of Suicide Assessment (C-CASA) will allow for valid and reliable comparisons between studies.

In other words, the consensus of mental health clinicians and researchers is that C-CASA makes a positive contribution to the field, one which will promote the reliability of suicide-related terminology and, ultimately, classification of suicide-related thoughts and behaviors.

**Criticisms of C-CASA**

At least one study explicitly criticized the C-CASA. Fedyszyn, Harris, Robinson and Paxton (2012) stated:

One limitation of the C-CASA is the lack of clear operational criteria of the key constructs of self-injurious behavior and suicidal intent. The C-CASA does not seem to provide instructions on the decision-making process, or on how to handle ambiguous, incomplete, or contradictory information, although some attempt has been made to establish guidelines for inferring suicidal intent in cases where explicit statement of such is missing... the biggest limitation... is [its] use of suicide risk factors as a way of inferring suicidal intent in narratives in which explicit statements of intentionality are missing. (p. 152)

The article’s most ardent criticism is that the C-CASA used suicide risk factors to help infer suicidal intent when circumstances did not provide explicit evidence of such intent.
Another criticism of C-CASA concerns the fact that it is retrospective, rather than prospective, in nature. Dr. Stein, in an expert panel commentary on defining suicidal risk (Fawcett, Baldessarini, Coryell, Silverman, & Stein, 2009), stated:

There are a number of problems with the use of [C-CASA] in data analyses by the FDA. The data were not prospectively collected to assess suicidal intent or planning, and the term suicidality served as a composite variable that does not appear to have much validity. (p. 783)

Dr. Stein’s first point, about prospective and retrospective reporting, is unique to the way that drug companies report adverse events. Pharmaceutical manufacturers report adverse events themselves, and later the FDA (and investigators at Columbia) performed keyword searches on those reported narratives to identify the ones with suicide-related events. While this is relevant to the FDA’s conclusions about suicide risk for medications, it is not relevant to nomenclature issues, which are about terms and definitions rather than statistically-supported conclusions about adverse events with the use of drugs.

On the other hand, Dr. Stein’s second point, concerning the validity of the term “suicidality,” is highly relevant to suicide-related nomenclature. Dr. Silverman expanded upon this point in the same article (Fawcett et al., 2009):

I do not like to use the term suicidality, because it does not convey specific clinical information (e.g., whether the patient has experienced suicidal ideation, has made an attempt, or has been exposed to suicidal behaviors). Thus, the term is too broad. (p. 782)

In the FDA analyses, “suicidality” referred to any of the four categories “completed suicide,” “suicide attempt,” “preparatory acts toward imminent suicidal behavior,” or “suicidal ideation,” so Dr. Silverman’s point about it being a broad term seems appropriate (Meyer et al., 2010). In fact, Meyer et al. advocate that “suicidality” no longer be used because of its ambiguity. The question about whether it is too broad or ambiguous, though, is one that this dissertation will attempt to contribute data regarding, by looking at clinicians’ perceptions of the clinical utility of the term “suicidality.” Dr. Silverman did agree that the term “has been widely used since the FDA analyses” (Fawcett et al., 2009, p. 782) and thus it seems important to assess clinicians’ perception of its clinical utility. However, it also seemed important to point out that
most of the collection of 25 representatives of “academia, government, and industry” at the “consensus conference” represented by the publication of Meyer et al. (2010), agreed that suicidality “should be abandoned as a term.” (p. 1042) because of its ambiguity, preferring instead to use the terms “suicidal ideation,” “suicidal behavior,” and “suicide”; Posner and Oquendo, two of the authors of C-CASA (Posner, 2007), were representatives contributing to Meyer et al. (2010).

**Literature Review of the Silverman Nomenclature**

Like the C-CASA, there is a substantial literature concerning the Silverman nomenclature. An overview of Silverman nomenclature will be presented first. Next, articles that used significant portions of the Silverman nomenclature will be reviewed. Articles will then be reviewed in which the Silverman nomenclature was cited for only one or two specific definitions. Finally, the few criticisms of the Silverman nomenclature will be addressed.

**Overview of the Silverman Nomenclature**

In contrast to C-CASA – which was a practical classification system – the Silverman nomenclature was first and foremost a theoretical undertaking. Silverman et al. (2007a) was responding to the conceptual confusion and variety of terminology that had existed in the field of suicidology since its inception (O’Carroll et al., 1996). Because of this confusion, Silverman et al. (2007b), which defined the Silverman nomenclature, left to future research the application of the nomenclature to clinical research. Unfortunately, since its publication, no research has looked at the reliability of the terms and definitions contained within the Silverman nomenclature. The Silverman nomenclature has thus been less advanced than C-CASA, which has had research done on its interrater reliability (Posner et al., 2007).

This does not mean, however, that the Silverman nomenclature has not been well received by the research community. The Silverman nomenclature was well received and influential: it has been cited approvingly in an astonishingly wide variety of types of suicide-related research. For instance, the Silverman nomenclature was commended in the study of substance abuse and suicidal behavior (e.g., studies of suicide and alcohol (Pompili et al., 2010), suicide attempts and substance abuse (Britton &
Connor, 2010), and addictive substances and unnatural death (Brådvik, Berglund, Frank, Lindgren, & Löwenhielm, 2009). It was likewise cited in theoretical work on suicide and self-injury, such as Joiner’s interpersonal-psychological theory of suicidal behavior (Van Orden et al., 2010), attachment theory and self-directed aggression (Gormley & McNiel, 2010), attachment styles and suicide-related behavior (Stepp et al., 2008), and in the identification of homogenous subgroups of suicidal patients (Mission et al., 2009).

The Silverman nomenclature was cited approvingly in studies of adolescents and children, such as in studies distinguishing subgroups of adolescents who self-harm (Hargus, Hawton, & Rodham, 2009), in research examining childhood trauma and risk for suicide attempts in prison among women (Clements-Nolle, Wolden, & Bargmann-Losche, 2009), in studies of bullying and suicide attempts among adolescents in custody (Kiriakidis, 2008), and in work that looked at trauma and suicide among Mexico City adolescents (Borges et al., 2008). The Silverman nomenclature was praised in self-injury research, such as a study of psychopathy and its relation to suicide and self-injurious behavior (Swogger, Conner, Meldrum, & Caine, 2009), research on self-harm and adolescents (Muehlenkamp, Cowles, & Gutierrez, 2010), and the introduction to the special section on suicide and nonsuicidal self-injury in the Journal of Consulting and Clinical Psychology (Prinstein, 2008). The Silverman nomenclature also found support in studies of disparate other suicide-related topics such as mid-life suicide (Hu, Wilcox, Wissow, & Baker, 2008), suicide and Parkinson’s disease (Kostić et al., 2010; Nazem et al., 2008), the neurobiology of suicide attempts among depressed patients (Pompili et al., 2008), and a review of “intent” in the suicide literature (Hasley et al., 2008).

As with the C-CASA nomenclature, the extant research on the Silverman nomenclature will be reviewed below. The reader will notice that, perhaps due to its theoretical rather than pragmatic origins, there are far fewer studies that adopted the Silverman nomenclature in its entirety. Rather, particular empirical studies used the Silverman nomenclature for specific definitions according to the studies’ specific needs, such as the definition of suicide attempt.
Studies That Adopted Significant Portions of the Silverman Nomenclature

At least six studies adopted the Silverman nomenclature for two or more definitions of suicide-related thoughts or behavior. The most comprehensive, which adopted the Silverman nomenclature in its entirety, was a recent Taiwanese conceptual analysis of suicidal behavior (Sun, 2011). That study focused on explaining the concept of suicide to Taiwanese nurses because of an important policy change by the Taiwanese government. Following a significant increase in suicides in Taiwan (the rate nearly tripled in the last fifteen years), the Taiwanese Department of Health instituted a policy in which public health nurses were sent to suicidal patients’ homes at least three times to assess suicidal risk and to provide support (Sun, 2011). The reason it is crucial for Taiwanese nurses to understand the conceptual basis of suicide was stated by Sun (2001): “understanding is currently lacking, in large part because public health nurses do not undertake continuing education and training on suicide and the process of suicide care.” (p. 458). To remedy this lack of understanding, Sun identified uses of the concept of suicidal behavior with reference to Silverman et al. (2007b): “These subtle differences lead to three classifications of suicidal behavior in the literature: (1) suicide-related ideation; (2) suicide-related communication (suicide threats and suicide plans), and (3) suicide-related behavior (self-harm, suicide attempts, and suicide); Silverman [et al., 2007b].” (Sun, 2011, p. 459). This is a perfect summary of the Silverman nomenclature and demonstrates the nomenclature’s reach, even internationally.

Another international study (this time in Canada) was that of Bergmans and Links (2009), who researched group interventions to reduce risk factors for suicide-related behavior. Bergmans and Links (2009) explained their use of the Silverman nomenclature as follows:

For the purposes of this article, the nomenclature of Silverman et al. [2007b] will be used, with particular focus on suicide-related behavior, defined as “a self-inflicted, potentially injurious behavior for which there is evidence (either explicit or implicit) either that (a) the person wished to use the appearance of intending to kill himself/herself in order to attain some other end; or (b) the person intended at some undetermined or some known degree to kill himself/herself. Suicide-related behaviors can result in no injuries, injuries, or death.” (p. 18)
In other words, Bergmans and Links (2009) adopted the Silverman nomenclature for suicide-related behavior, especially as it concerned suicide attempts and undetermined suicide-related behavior. The authors found that a 20-week group intervention resulted in significant pre- and post-group decreases in suicide risk factors such as hopelessness, satisfaction with life, depression, alexithymia, and impulsivity.

Another study that adopted large sections of the Silverman nomenclature was Borges et al. (2008), which studied trauma and suicide-related outcomes in about 3,000 adolescents in Mexico City. As with the article just discussed, even the title of the article included the term “suicide-related,” which is specific to the Silverman nomenclature. Furthermore, “in tandem with current nomenclature (Silverman et al., 2007a,” the study defined “suicide ideation,” “suicide plan,” and “suicide attempts” consistently with the Silverman nomenclature.

Borges et al.’s (2008) use of the Silverman nomenclature’s term, “suicide plan,” was noteworthy because the Silverman nomenclature is the only one that defined and used the term “suicide plan.” Borges et al. defined “suicide plan” as referring “to instances in which one takes the additional step of formulating the method and program of action through which one expects to carry out the [suicide] attempt.” (2008, p. 655). This definition used the language of the full definition of “suicide plan” in Silverman et al. (2007b), which is: “A proposed method of carrying out a design that will lead to a potentially self-injurious outcome; a systematic formulation of a program of action that has the potential for resulting in self-injury.” (p. 268). The Borges et al. (2008) definition was adapted to be specific to suicide attempts, but one can see its reliance on the “program of action” and “method” language from the Silverman definition. Borges et al. ultimately found that experiencing traumatic events increased the odds of adolescents having suicidal ideation by more than three times, increased the odds of having a plan more than five times, and increased the odds of attempting suicide more than 6.5 times (compared to community prevalence estimates).
In a study of the effects of combat on risky and self-destructive behavior, Thomsen, Stander, McWhorter, Rabenhorst, and Milner (2011) cited Silverman et al. (2007b) for the proposition that self-harm behaviors “may be subsumed under the rubric of ‘self-injurious thoughts and behaviors’ (Silverman et al., 2007)” (p. 1321), which is an accurate description of the superset containing the Silverman nomenclature. In this context, a “superset” is contrasted to a “subset”; in other words, a suicide-related nomenclature contains subsets such as suicidal thoughts and suicidal behaviors, whereas a superset’s categories (such as self-injurious thoughts and behaviors) contain all of the terms and definitions of a suicide-related nomenclature, as well as other terms like thrill-seeking behaviors (like bungee jumping or skydiving). Another study, Muehlenkamp, Cowles, and Gutierrez (2010), adopted the Silverman nomenclature in large part (through generally accurate paraphrases), though it did not follow it entirely with respect to what the Silverman nomenclature calls “self-harm”:

The following definitions based on the Silverman et al. (2007) taxonomy were used: self-harm is used as a general descriptor of behaviors both with and without intent to die; nonsuicidal self-injury (NSSI) is an act resulting in immediate tissue damage in which there was no conscious intent to die. (p. 236)

In fact, however, self-harm in the Silverman nomenclature means only those behaviors for which there is implicit or explicit evidence that the person had no intent to die at all; the term “nonsuicidal self-injury” was not used in the Silverman nomenclature.

Finally, another study that incorporated significant portions of the Silverman nomenclature was Nazem et al. (2008), which analyzed suicidal ideation, death ideation, and suicide attempts in those with Parkinson’s disease. Citing the Silverman nomenclature, Nazem et al. (2008) stated, “Suicide-related thoughts and behaviors may be categorized as suicide-related ideation, suicide attempts, and completed suicide.” (p. 1573). This is the only study which used the term “suicide-related ideation” from the Silverman nomenclature – the vast majority of research used the simpler term “suicidal ideation.” The study found that 28 percent of those with Parkinson’s disease had current death ideation (defined as the wish to die without suicidal intent), another 11 percent had current suicidal ideation, and 4 percent had attempted suicide at
The majority of citations to the Silverman nomenclature – excepting those which simply referred to Silverman et al. (2007a,b) for the proposition that terms and definitions in suicidology need improvement – had to do with citing the Silverman nomenclature for one particular definition. For example, like many other authors, Preti (2012) quoted Silverman (2007b) only for the definition of “suicide attempt”; likewise, Ghahramanlou-Holloway, Cox, and Greene (2012) correctly quoted Silverman et al.’s (2007b) definition of a suicide attempt: “A suicide attempt is defined as a ‘self-inflicted, potentially injurious behavior with a nonfatal outcome for which there is evidence (either explicit or implicit) of intent to die’ (Silverman et al., 2007, p. 273).” (p. 234). Other studies that quoted Silverman et al. (2007b) for that same quotation were Bohnert, Roeder, and Ilgen (2011); Nenadic-Sviglin et al. (2011); and Smith, Cukrowicz, Poindexter, Hobson, and Cohen (2010).

Some authors defined suicide attempt in ways that were consistent with Silverman et al. (2007b), though not identical to it: for example, “A suicide attempt was defined as a nonfatal, self-directed, potentially injurious behavior with any intent to die as a result of the behavior, regardless of outcome (cf. Crosby, Ortega, & Melanson, 2011; Silverman et al., 2007).” (Bryan & Rudd, 2012, p. 844). Likewise, Blasco-Fontecilla et al. (2011) utilized a definition of suicide attempt that was close to that of Silverman et al. (2007b), even using a quotation that captured all of the elements except for the idea that the behavior is potentially injurious: “Suicide attempts were defined as ‘a self-destructive behavior with intent to end one’s life independent of resulting damage’ (O’Carroll et al., 1996; Silverman et al., 2007a,b).” (p. 239).

Some authors stated that they used Silverman et al. (2007b) for the definition of suicide attempt, though they missed some element of it or paraphrased it incorrectly. For example, Vishnuvardhan and Saddichha wrote, “A suicide attempt refers to engagement in a potentially self-injurious behavior in which there is at least some intent to die.” (2012, p. 410). While this lacked the “implicit or explicit evidence”
clause, it was not nearly as poor a paraphrase as that of Monnin et al. (2012), who wrote, “A suicide attempt was defined as self-inflicted and potentially dangerous behavior for which the intent to die has been proven, according to Silvermann [sic] et al. (Silverman et al., 2007).” (p. 36). In fact, suicidal intent need not be “proven”; rather, there need only exist some evidence (either explicit or implicit) of intent to die. Likewise, Pompili et al. (2011) defined suicide attempt: “A suicide attempt was defined as a self-destructive act with some degree of intent to end one’s life and some identifiable injuries.” (p. 301). However, the Silverman nomenclature did not require injuries; rather, it differentiated between Type I (without injuries) and Type II (with injuries), both of which are suicide attempts.

Mazza, Catelano, Abbott, and Haggerty’s (2011) article on suicide attempts in youth was noteworthy in that it referred to Silverman et al.’s (2007b) two types of suicide attempt. As described above, a Type I attempt is one without injuries, whereas a Type II attempt is one with injuries. Mazza et al. noted that their study did not differentiate between Type I and Type II suicide attempts; however, it was still unusual in that it cited Silverman et al. for those two types of attempts. Another study that did use Type II suicide attempts only (not Type I) was Pompili et al. (2011). Finally, Zayas, Gulbas, Fedoravicius, and Cabassa (2010) used the Silverman nomenclature in their study of young Latinas and did in fact differentiate between Type I and Type II suicide attempts.

**Criticim of the Silverman Nomenclature**

There has been very little criticism of any part of the Silverman nomenclature. Only one study explicitly criticized the Silverman nomenclature: Kleespies et al. (2011) declined to use the Silverman nomenclature in their study of self-injurious and suicidal behaviors among veterans. Instead, they used a system in which self-injurious behaviors are the superset, divided into two types: suicidal (including “a gradation of intent to die”) and nonsuicidal (“with no intent to die”) (Kleespies et al., p. 238). They contrasted this with the Silverman nomenclature, which subsumed self-injurious behaviors under the superset of suicide-related behaviors “and intentional self-injurious behaviors with no suicidal intent are regarded as suicide-related.” (Kleespies et al., p. 238). However, other studies cited Silverman et al.
(2007b) approvingly for the Silverman nomenclature’s use of suicidal intent to distinguish between non-suicidal self-injury and suicidal self-injury (Andover & Gibb, 2010). Likewise, Joiner’s interpersonal theory of suicide – which addresses self-injurious behaviors as part of the acquired capability to kill oneself – explicitly adopted the Silverman nomenclature as its terminological and definitional basis (Van Orden et al., 2010; Smith & Cukrowicz, 2010).

Of course, any studies that elected to use the C-CASA or another nomenclature rather than the Silverman nomenclature were, in some sense, criticizing the Silverman nomenclature implicitly. Despite its strengths, the Silverman nomenclature is not the only nomenclature and classification system currently available. A review of the epidemiology of suicide listed Silverman, Berman, Sanddal, O’Carroll, and Joiner (2007b) as part of a “consensus” of recent research related to suicide nomenclature (Nock et al., 2008); however, alternative systems of classification were also included among the works cited, specifically the C-CASA, which was discussed above (Posner et al., 2007). Furthermore, the C-CASA is not the only competing suicide nomenclature: De Leo et al. (2006) offered their own definitions and proposed nomenclature building on both the O’Carroll nomenclature and the WHO/EURO Multicentre Study on Suicidal Behavior (Bille-Brahe, Kerkhof, De Leo, & Schmidtke, 2004). Lester (2009) proposed another suicide nomenclature that paralleled legal ideas regarding intent, premeditation, and provocation for murder: in his system, first-, second-, and third-degree suicide are analogous to first-, second-, and third-degree murder. First-degree suicide would require that the person have intent and premeditation; second-degree suicide likewise would need suicidal intent but the death would not be premeditated (e.g., it could have been impulsive); while in third-degree suicide a person would have intended to harm himself or herself but would have lacked the intent to die.

Utilization of the Silverman nomenclature has also not been universal within the realm of self-injurious behavior. For example, O’Connor, Rasmussen, and Hawton (2010) used the Child and Adolescent Self-harm in Europe (“CASE”) criteria (Hawton, Rodham, & Evans, 2006) for the definition of self-harm instead of the definition proposed by the Silverman nomenclature; O’Connor, Rasmussen, and Hawton
(2009) likewise mentioned the Silverman nomenclature while declining to adopt it with respect to deliberate self-harm. Shemesh et al. (2009) noted that the practical limitations of the assessment instruments they used did not permit them to make use of the Silverman nomenclature. One of the leading scholars of self-injury research disagreed with the decision of the Silverman nomenclature to not use the term "suicide gesture" (Nock, 2008). In contrast, Heilbron, Compton, Daniel, and Goldston (2010) argued that "suicide gesture" is a historically dismissive term and that nomenclatures could incorporate its critical content — that is, non-suicidal, self-injurious behavior that is intended to look like a suicide attempt to others — without using the loaded term "suicide gesture." Finally, there are those who noted that research is ongoing and the jury is still out regarding the best terminology to distinguish between suicide and self-harm behaviors (Plener, Libal, Keller, Fergert, & Muehlenkamp, 2009).

However, the Silverman nomenclature has been seen in generally positive terms by the research community. Wasserman et al. (2012) referred to the Silverman nomenclature as "a meritorious attempt to revise the nomenclature for the study of suicide..." King, O'Mara, Hayward, and Cunningham (2009), in their study of emergency department suicide risk screening of adolescents, wrote that the Silverman nomenclature was "a widely accepted taxonomy of suicidal behavior set forth by O'Carroll et al. [1996] and Silverman et al. [2007a,b]." (p. 1236). In fact, the research community has showed its respect for the Silverman nomenclature by citing Silverman et al. (2007a,b) approximately 150 times in the literature in a five-year span, according to the Social Science Citation Index (Thomas Reuters, 2012). This large research influence, coupled with the absence of an interrater reliability study on the Silverman nomenclature, highlights a need for this dissertation to collect interrater reliability data for the Silverman nomenclature.

**Collecting Evidence Regarding Suicide-Related Nomenclatures**

Moving on from literature reviews of the two major current suicide-related nomenclatures, the types of evidence that are relevant to evaluating suicide-related nomenclatures will now be discussed. First the history of the type of evidence that has supported suicide-related nomenclatures will be described; this history is predominantly the history of the outcomes of expert workgroups. Next, the types
of evidence that have been utilized in the medical field will be contrasted with that of the field of suicidology. Specifically, interrater reliability has been the predominant evidence collected for medical nomenclatures. As the primary mental-health-related nomenclature, the DSM likewise is supported by interrater reliability evidence. Next, this section will argue that the DSM can also serve as a blueprint for the types of evidence that can be collected concerning the clinical utility of suicide-related nomenclatures. Finally, the three types of clinical utility evidence will be reviewed: ease of usage, communication, and treatment planning.

Nomenclature Evidence in Suicidology Has Predominantly Been Collected Through Workgroups

In the field of suicidology, the types of evidence that have typically been collected regarding suicide nomenclatures have been quite limited. The largest type of evidence has been informal: expert consensus reached in workgroups. For example, in 1970, the Committee on Nomenclature and Classification, chaired by Aaron Beck, considered suicide-related definitions and nomenclature from the 1950s and 1960s; many terms were examined, including "suicide gesture," "self-damaging behavior," "ambivalent suicide attempt," "serious suicide attempt," "suicide thoughts or ideas," "suicidal preoccupation and ruminations," as well as the three main categories the committee decided upon, namely, "completed suicides," "suicide attempts," and "suicidal ideas." (Pokorny, 1974, pp. 31-36).

After this seminal workgroup in the field of suicidology, it took another decade before a second workgroup was convened. In order to address the problem of whether an ambiguous set of circumstances surrounding a death should be called a "suicide," in the mid-1980s, the Centers for Disease Control (CDC) brought together a multidisciplinary workgroup charged with the task of creating a unified set of criteria that would guide coroners and medical examiners. It is the coroner’s or the medical examiner’s job to make the determination about whether a given death should be classified as a suicide death. Coroners and medical examiners are trained differently in different states; the former are often appointed and both receive relatively little training in determining suicidal intent (Jobes, Berman, & Josselson, 1987). The CDC workgroup created 11 examples of explicit or implicit evidence that might assist the coroner or medical
examiner in determining whether or not a particular death was in fact suicide. The CDC workgroup also noted that there was a definition of suicide "implicit in these criteria": "Death arising from an act inflicted upon oneself with the intent to kill oneself." (Rosenberg et al., 1988, p. 1451). Until the definition of suicide moves from implicit to explicit, with consistent training for coroners and medical examiners on the subtleties of determining suicidal intent, official suicide statistics will remain flawed (De Leo et al., 2006; Jobes et al., 1987; Rosenberg et al., 1988). A standardized, reliable nomenclature that, among other things, carefully and thoroughly defines the term “suicide,” is necessary to achieve that end (De Leo et al., 2006; Silverman et al., 2007a).

Again, it took another decade before another set of workgroups was convened. The field of suicidology took great steps forward in selecting consensus terminology during two workshops held in 1994. After informal discussion for several years, in April of 1994 a workshop was convened by the American Association of Suicidology with the purpose of developing a common nomenclature. This workshop was followed by another in November of 1994 in Washington, D.C., sponsored by the National Institute of Mental Health. These workshops culminated in the publication of a seminal article on suicide nomenclature: "Beyond the Tower of Babel: A nomenclature for suicidology." (O’Carroll et al., 1996). This article (hereinafter called “the O’Carroll nomenclature”) was the intellectual predecessor of the Silverman nomenclature; Silverman was one of the authors of the O’Carroll nomenclature and the Silverman nomenclature was a deliberate revision of the O’Carroll nomenclature.

The O’Carroll et al. (1996) nomenclature divided suicide-related behavior into two categories: instrumental suicide-related behaviors and suicidal acts. O’Carroll et al. defined instrumental suicide-related behavior as follows:

Potentially self-injurious behavior for which there is evidence (either explicit or implicit) that (a) the person did not intend to kill himself/herself (i.e., had zero intent to die) and (b) the person wished to use the appearance of intending to kill himself/herself in order to attain some other end (e.g., to seek help, to punish others, to receive attention). (p. 247)

In other words, instrumental suicide-related behavior included acts in which the person had no suicidal intent but in which he or she used the appearance of suicidal intent as an instrument (hence the word...
“instrumental suicide-related behavior”) to accomplish something else. Instrumental suicide-related behavior was subdivided into three further categories: with injuries, without injuries, and with fatal outcome. Likewise, suicidal acts were divided into suicide attempts (either with or without injuries) and completed suicide. Definitions were provided for each of the key terms in this nomenclature: suicide, suicide attempt with injuries, suicide attempts, suicidal act, instrumental suicide related behavior, suicide related behavior, suicide threat, and suicidal ideation. In spite of the fact that O’Carroll et al. (1996) has been cited about 300 times according to the Social Science Citation Index (Thomas Reuters, 2012), no articles presented any data about the reliability of the O’Carroll nomenclature.

Though the O’Carroll nomenclature made explicit the confusion in the literature and in clinical practice that surrounded suicide-related nomenclature – as well as offering a nomenclature that appeared intuitive and thoughtful – it was not a perfect system. Silverman et al. (2007a) found that clinicians were hesitant to use one of the O’Carroll nomenclature’s key terms: "instrumental suicide-related behavior" was considered too much of “a ‘mouthful’ and easily misunderstood, and would be difficult to insert into the existing vocabulary of suicidology." (Silverman, Berman, Sanddal, O’Carroll, & Joiner, 2007a, p. 257). As a result, Silverman et al. (2007a) reported that a nomenclature workgroup was formed in the 19th geographic area of the Veterans Integrated Service Network (“VISN 19,” which covers roughly the Rocky Mountains area), and was tasked with attempting to improve the O’Carroll nomenclature:

[The] Denver Veterans Administration VISN 19 Mental Illness Research, Education, and Clinical Care (MIRECC) nomenclature workgroup reviewed the critiques and recommendations made in response to the O’Carroll et al. paper, and proceeded to address as many as possible in the revision. (p. 259).

The revision of the O’Carroll nomenclature was presented in the second article of a two-article series in the journal Suicide and Life-Threatening Behavior in 2007 (Silverman et al., 2007b), and has been referred to in this dissertation as “the Silverman nomenclature.”

As one can see from the history laid out above, a series of expert workgroups and workshops created the major nomenclatures in the field of suicidology. Other work in creating nomenclatures has
typically been done by individuals or authors who believed that they had good ideas for nomenclature advancement (Lester, 2009; De Leo et al., 2006). Suicidology is not alone in this methodology; for example, the Current Procedural Terminology® (“CPT”) utilized in the medical field has been the result of workgroup discussions as well. However, these workgroups have been confidential affairs convened by the American Medical Association, which holds the copyright on CPT codes used in medical billing (APA, 2012). This secrecy has made it impossible for researchers to reproduce the process involved in the creation of that nomenclature.

**Medical Nomenclature Evidence Has Predominantly Been Collected Via Interrater Reliability**

In contrast, the medical field has advanced beyond the field of suicidology in that it has made a habit of publishing reliability statistics regarding nomenclatures within specific medical domains. Studies of nomenclatures in medicine predominantly focus on interrater reliability of terms and definitions, though some also focus on test-retest reliability. For example, Hirsch (2011) assessed interrater reliability of a standardized nomenclature for describing electrographic (EEG) patterns, developed by the American Clinical Neurophysiology Society over a six-year period; test-retest reliability was not discussed. Similar nomenclature studies that assessed interrater but not test-retest reliability in the medical field include Offiah et al. (2010) (nomenclature for classifying congenital spinal defects); Baker, Ben-Tovim, Buchter, Esterman, and McLaughlin (2007) (nomenclature for classifying voice disorders); Chau, Desai, Geogalas, and Harries (2005) (nomenclature for benign laryngeal pathology); and Redline et al. (2004) (nomenclature for fetal vascular obstruction lesions).

When the broader literature on the reliability of medical nomenclatures was examined, all nine of the relevant publications calculated interrater reliability, while only one calculated test-retest reliability but not interrater reliability (Arana et al., 2011; Bono et al., 2011; Hirsch, 2011; Offiah et al., 2010; Baker, Ben-Tovim, Buchter, Esterman, & McLaughlin, 2007; Hasin, Hatzenbueler, Keyes, & Ogburn, 2006; Chau, Desai, Geogalas, & Harries, 2005; Redline et al., 2004; Hruban et al., 2001; Demetris et al., 1995). However, four of the nine studies that looked at interrater reliability also looked at test-retest reliability.
(Arana et al., 2011; Bono et al., 2011; Hruban et al., 2001; Demetris et al., 1995). In the medical field, therefore, researchers have consistently measured the interrater reliability of nomenclatures, while about half the time they have also examined test-retest reliability.

However, there were some examples of assessment of both interrater and test-retest reliability. Bono et al. (2011) had surgeons apply a nomenclature for describing spinal injuries to radiographic images at time 1 and time 2; the authors calculated both interrater and intrarater kappas and percent agreement statistics. Arana et al. (2011) likewise obtained interrater and intrarater reliabilities for magnetic resonance imaging of lumbar disc contours, in order to compare two competing spinal nomenclatures. Hruban et al. (2001) measured both types of reliability for a new nomenclature concerning pancreatic duct lesions. Demetris et al. (1995) tested a nomenclature for liver transplant rejection using both interrater and intrarater reliability.

Finally, there were examples of nomenclature studies that involved only test-retest rather than interrater reliability. Hasin, Hatzenbueler, Keyes, and Ogburn (2006) conducted a literature review of studies examining reliability and validity evidence for DSM-IV versus ICD-10 substance dependence and substance abuse definitions; they reported several studies involving test-retest reliability but did not report any interrater reliability studies.

In studies of the reliability of DSM categories, similar patterns were seen to that of the medical nomenclature described above. Specifically, seven of fifteen studies since 2000 looked at interrater reliability alone (Brown, Di Nardo, Lehman, & Campbell, 2001; Gutkind, Ventura, Barr, Shaner, Green, & Mintz, 2001; Jakobsen et al., 2005; Jakobsen et al., 2008; Jovanovic, GaSic, Ivkovic, Milovanovic, & Damjanovic, 2008; Shore, Savin, Orton, Beals, & Manson, 2007); six examined both interrater reliability and test-retest reliability (Grant, Steinberg, Kim, Rounsaville, & Potenza, 2004; Pierucci-Lagha et al., 2007; Ries, Demirsoy, Russo, Barrett, & Roy-Byrne, 2001; Schneider et al., 2004; Schneider et al., 2005; Zanarini & Frankenburg, 2001; Zanarini et al., 2000), while only two studies (Cottler, Leung, & Abdallah, 2009; Mennes, Abdallah, & Cottler, 2009) examined test-retest reliability without also studying interrater
reliability. In other words, studies of reliability of DSM categories showed the same pattern that other medical categories did: interrater reliability was almost always assessed, while only a few studies looked at test-retest reliability without also examining interrater reliability.

In the medical field, therefore, nomenclatures have been assessed utilizing reliability statistics. Typically, reliability was measured using interrater kappas, though some studies used test-retest scores in addition. This puts the medical field ahead of the field of suicidology in terms of its commitment to empirical reproducibility of its nomenclature. This dissertation will thus contribute to a needed area in suicidology by collecting interrater reliability data on the two major current suicide-related nomenclatures.

**The DSM Provides a Blueprint for Testing Suicide-Related Nomenclatures’ Reliability**

As discussed in the previous chapter, the DSM is a nomenclature that provides a blueprint for testing nomenclatures in the field of suicidology. Therefore, we can look to the ways in which prior and current versions of the DSM had their reliability tested in order to discover ways in which suicide-related nomenclatures’ reliability might be assessed. As discussed immediately above, the primary way in which DSM categories had their reliability tested was through examination of interrater reliability. As such, test-retest reliability of suicide-related nomenclatures will be left for future studies to assess.

Several articles described testing the interrater reliability of DSM categories in ways that are instructive regarding testing suicide-related nomenclatures. Zimmerman (2011) laid out the history of DSM reliability for personality disorders, noting the DSM-II displayed poor interrater reliability (mean kappa for personality disorders was .44, based on a prototype-matching process) while DSM-III showed improved interrater reliability (mean kappa for personality disorders was .61, based on explicit definitional criteria, which permitted standardized diagnostic interviews to be developed). Using explicit definitional criteria (such as those in suicide-related nomenclatures to be studied in this dissertation) improved interrater reliability significantly; Zimmerman (2011) argued that such a result had implications for the reliability of proposed prototype matching for personality disorders in DSM-5: “The available
evidence therefore suggests that a return to a DSM-II diagnostic approach based on matching patients to prototypic descriptions will result in a return to the problem of poor reliability faced in the pre-DSM-III era.” (p. 214). Applied to suicide-related nomenclatures, the above suggests that explicit definitional criteria – such as those created by all two of the major current suicide-related nomenclatures – are very important to creating nomenclatures that have adequate interrater reliability.

Interrater reliability was also assessed in the DSM-IV. Klin, Lang, Cicchetti, and Volkmar (2000), four researchers from the Yale Child Study Center, looked at the results of the DSM-IV autism field trial. That study computed interrater reliability among a large number (N = 83) of raters. Klin et al. concerned the diagnosis of autism, which had previously assessed interrater reliability most often among autism experts. However, Klin et al. argued that “diagnostic assessments are made in the larger clinical community probably as often as by the smaller number of autism ‘experts.’ Hence the importance of expanding the empirical verification of reliability... into the community of less experienced clinicians.” (pp. 163-164). This required a larger number of raters for interrater reliability. Likewise, judgments about suicide risk are regularly made in the larger clinical community much more often than by the smaller number of suicide experts. As such, it is just as important for this dissertation to examine reliability among clinicians in the field (even less experienced ones) as it was for previous research to collect workgroup evidence from among suicide experts. Like Klin et al., this will require a larger number of raters.

Klin et al. (2000) also emphasized the DSM’s status as a nomenclature, specifically by pointing out that definitions are key to judgments about the reliability of the DSM-IV: “whether or not one utilizes... diagnostic instruments, the diagnostic assignment still depends on the adoption of consensual definitions as operationalized by the DSM-IV...” (p. 163). Likewise, reproducible judgments about suicide risk depend upon the adoption of consensual definitions as operationalized by current suicide-related nomenclatures.

The way in which judgments are reproducible in research depends on the methodology chosen by the researcher. Fogelson, Nuechterlein, Asarnow, Subotnik, and Talovic (1991) studied interrater
reliability of Axis II for DSM-III-R; the authors laid out the differences between “test-retest” and “same interview” methods of examining interrater reliability: “In the test-retest paradigm, the same subject is interviewed by one interviewer, and again later by a second interviewer. This procedure introduces information variance due to differing interviewing styles of the diagnosticians.” (p. 56). This information variance lowers reliability scores improperly. In contrast, “the same interview paradigm, where the subject is interviewed by one interviewer while the second interviewer silently observes and makes independent ratings, isolates the variance due to differences in diagnostic decision making...” (Fogelson et al., 1991, p. 56). This highlights the importance of limiting information variance. While conducting interviews of suicidal patients would be both impractical and ethically difficult, information variance can be limited through vignette research, in which the clinician participants rate the same vignettes, which necessarily contain the same information. Since both C-CASA and the SDVCS also used narrative vignettes, this would be consistent with prior research on suicide-related nomenclatures.

Clinical Utility and the DSM

The DSM also provides a blueprint for the assessment of clinical utility of suicide-related nomenclatures. Providing clinical utility was one of the foremost goals of the architects of DSM-IV: the introduction to the DSM-IV stated, “Our highest priority has been to provide a helpful guide to clinical practice.” (APA, 2000, p. xxiii). While this “helpful guide to clinical practice” language was an intuitive description of what was meant by clinical utility, Mullins-Sweatt and Widiger (2009), in their article, “Clinical Utility and DSM-V,” laid out the three components of clinical utility more precisely: “three primary components are the central matters of clinical utility: ease of usage, communication, and treatment planning.” (p. 303). Each of these three components will be explored as they relate to both the DSM and to suicide-related nomenclatures.

Ease of Usage in Clinical Practice

The DSM-III and DSM-III-R were great steps forward in terms of reliability but posed problems in terms of ease of usage in clinical practice, primarily because they were written by researchers for
researchers (Kendler, 1990). Examples of diagnoses that posed problems for clinicians were Generalized Anxiety Disorder (which contained 22 criterion sets) and Somatization Disorder (which contained 35 criterion sets) (APA, 1987); these lengthy criterion sets might have been suitable for research settings in which researchers have more time to conduct interviews, but busy clinicians often could not afford that kind of interviewing time (Sweatt & Widiger, 2009).

Likewise, both O’Carroll et al. (1996) and Silverman et al. (2007a) commented on the importance of retaining terms that were already in widespread usage in the clinical community. One reason for this could be seen as stemming from ease of usage: it would be easier for clinicians to incorporate preexisting terms into their vocabulary than to try to learn an entirely new set of terms and definitions. Furthermore, Silverman et al. (2007a) stressed the importance of simplicity of terms, which also streamlines the ease of usage of a suicide-related nomenclature. Finally, Silverman et al. (2007a) reported receiving more than 100 communications (e.g., e-mails, phone calls) after the publication of the O’Carroll nomenclature in which people stated that several proposed terms (such as “instrumental suicide-related behavior” or “non-zero intent”) “were deemed by others to be too broad, too vague, or too unwieldy.” (p. 251). Unwieldy terms would most likely be problematic because they would impair ease of usage of the nomenclature.

**Communication as a Component of Clinical Utility**

Sweatt and Widiger (2009) wrote that “the primary purpose of an official diagnostic nomenclature is to provide a common language of communication.” (p. 306). Therefore, the DSM-I was introduced because of the “crippling confusion” engendered by the absence of a common language for clinical practice and researchers (Sweatt & Widiger, 2009, p. 306). This may remind us of the current state of suicidology, in which common terms and definitions do not exist in the clinical and research communities (O’Carroll et al., 1996; Silverman et al. 2007a).

Facilitating communication among researchers is an obvious goal of both DSM and suicide-related nomenclatures, since researchers need to have similar definitions of key constructs in order to
make research comparable across studies (Spitzer, First, Schedler, Westen, & Skodol, 2008). Likewise, for clinicians, it is important that they be able to communicate with each other (and insurance companies) in terms that everyone can understand. This is one of the debates surrounding the adoption of a dimensional classification by DSM-5 for personality disorders: will the shift from categorical to dimensional descriptions make it easier or harder for clinicians, researchers, and insurance companies to communicate with each other? (Mullins-Sweatt and Widiger, 2009). In one sense, categorical classifications are simpler: the person either has or does not have the category at issue. However, a dimensional description may be more accurate, thus facilitating more clear communication. Thus there can be tension between ease of usage and facilitating communication.

For suicidology, too, this tension has existed between the simplicity of a nomenclature and the clarity of its communication (Silverman et al., 2007a). For example, whereas “suicide-related ideations” might be more precise (thus increasing clarity of communication), “suicidal ideation” might promote easier usage. With that example, clinical practice suggests that ease of usage may trump precision of communication, since clinicians virtually all use “suicidal ideation” as opposed to “suicide-related ideations.”

Clarity of communication is not simply limited to communication between clinicians, however. The DSM-IV and the Five Factor Model of personality have both been rated for their utility of professional communication between clinicians, as well as communication with clients or other laypeople (such as family members or significant others of the client) (Mullins-Sweatt & Widiger, 2011). In fact, the utility of the DSM-IV and Five Factor Model did not differ significantly in terms of clinicians’ ratings of their ability to promote clarity of communication with other clinicians, but they did differ in terms of communicating information to clients (with the Five Factor Model rated as superior to the DSM) (Mullins-Sweatt & Widiger, 2011). Therefore, in assessing the perceived clinical utility of suicide-related nomenclatures, it would be beneficial to ask clinicians to rate both communication between clinicians and communication with clients or other laypersons.
Mullins-Sweatt and Widiger (2009) referred to changes in treatment outcomes as the “holy grail” of clinical utility (p. 308). Mullins-Sweatt and Widiger, however, stated that “it might be with respect to treatment planning that the DSM-IV-TR is most problematic.” (p. 308). In support of this proposition, Mullins-Sweatt and Widiger quoted the chair and vice-chair of the DSM-5, who commented on the lack of success of the DSM-IV in contributing to treatment planning: “With regard to treatment, lack of treatment specificity is the rule rather than the exception.” (p. 308, citing Kupfer et al., 2002, p. xviii). In exploring the reasons for this lack of treatment specificity, Mullins-Sweatt and Widiger pointed out that the authors of the DSM-III (and DSM-IV) intentionally tried to make the manual neutral in terms of theoretical orientation. The reason for this theoretical neutrality, stated Mullins-Sweatt and Widiger (2009), was that “a language that favors one particular perspective does not provide an equal playing field” for all theoretical orientations (p. 308). Until substantial evidence favors one or several theoretical orientations as more efficacious than others, skewing the diagnostic language towards any particular perspectives would not contribute to clinical utility.

Since evidence for particular treatments of suicidal thoughts and behaviors is sparse, it would likewise be premature to let a particular theoretical orientation determine the language of suicide-related nomenclatures. Instead, clinical utility for treatment might be guided by “consumer preference” studies, in which the preferences of clinicians for specific terms and definitions as they relate to treatment planning might be given weight in the choice to adopt particular suicide-related terms and definitions. This is akin to the “consumer preference” study of five different dimensional systems for personality disorder diagnosis, in which 400 randomly selected psychiatrists and psychologists rated the clinical utility of those five dimensional systems (Spitzer, First, Shedler, Westen, & Skodol, 2008). While the specific methods used in that consumer preference study will be detailed in the Method chapter of this dissertation, for present purposes it seems sufficient to note that the authors asked seven questions regarding perceived clinical utility for each of the five dimensional personality systems; those questions
encompassed the clinician’s global impression of usefulness and feasibility of the system, ease of usage, usefulness for communication to other clinicians, and usefulness in treatment planning. In other words, the study’s design was consistent with Mullins-Sweatt and Widiger’s tripartite framework for exploring clinical utility. The study found that the Shedler-Westen Assessment Procedure (SWAP) prototype matching system and the DSM-IV prototype matching system were the models rated the most clinically useful by clinicians.

Samuel and Widiger (2011) criticized Spitzer et al. (2008) on the grounds that clinicians’ preference for prototypal matching might be simply a function of ease of usage. They stated that “whereas the prototypal matching required very little time, the [Five Factor Model] ratings required a consideration of six pages of material. Thus, the results might simply reflect a preference by the clinicians for the easier method.” (p. 14). While this could be true, we have seen in the above discussion that ease of usage is one component of clinical utility (Mullins-Sweatt & Widiger, 2009), so this fact would not, on its face, prevent Spitzer et al.’s conclusions from being correct.

However, Mullins-Sweatt & Widiger (2011) were able to collect utility ratings using two one-page rating forms: one summarized both the 10 DSM-IV personality disorders in dimensional terms, and the other summarized the 30 facets of the Five Factor Model. With these rating forms of equivalent length – despite clinicians’ greater familiarity with the DSM-IV system – Mullins-Sweatt & Widiger found that the Five Factor Model was rated higher than the DSM-IV dimensional system on four of six components of clinical utility. Interestingly, the two components on which the systems did not differ were ease of usage and communication between clinicians. However, clinicians did rate the Five Factor Model as more useful in terms of communicating information to clients, for a description of the client’s personality problems, for giving useful information for treatment planning, and for yielding a global description of personality. Again, therefore, the DSM-IV model was found lacking in the area of assisting with treatment planning.

This study also points out the importance of creating summary forms for each system that are roughly equivalent in length, if possible. The DSM-IV prototypal matching system was found to be
superior to the Five Factor Model system when the latter’s summary form was six times the length of the former. In contrast, when both systems were summarized in forms of comparable length (i.e., one-page descriptions), the Five Factor Model was deemed superior. Therefore, in designing summaries of suicide-related nomenclatures, it would be important to create summaries that are comparable in length.

Summary of Literature Review

The literature surrounding the two major current suicide-related nomenclatures was summarized in this section. The C-CASA was developed as a practical classification tool to assist the FDA; upon its adoption as a requirement in clinical trials by the FDA, it has been used in hundreds of clinical trials that used the C-CASA. It has displayed good interrater reliability, though there is some variability in terms of its categories, some of which are quite reliable and others of which have problematic reliability. The C-CASA has had the most practical influence of any of the two nomenclatures.

In contrast, the Silverman nomenclature was designed as a theoretical nomenclature rather than a classification system. The most practical application of the Silverman nomenclature has been many studies’ adoption of its definition of “suicide attempt.” As a nomenclature, Silverman et al. (2007a,b) has been widely cited – even internationally – in the academic research literature on suicide, especially for the proposition that there is great conceptual confusion and variety of terms used in the nomenclature of suicidology. No reliability data have yet been collected for the Silverman nomenclature.

Suicide-related nomenclature evidence has primarily been collected through expert workgroups. This is unlike the medical field and the DSM, both of which have typically collected interrater reliability data on nomenclatures. Interrater reliability is the dominant method of assessing reliability and allows for reproducibility of evidence, unlike expert workgroups. There is a problem of information variance if multiple interviewers are used, but utilization of vignettes solves this problem, as well as being consistent with prior research (on the C-CASA and the SDVCS).

Since it is the dominant mental health nomenclature, the DSM can be used as a blueprint for testing suicide-related nomenclatures. With respect to research on personality disorders of the DSM,
three components of clinical utility have been described: ease of usage, communication, and treatment planning. Communication included two parts: communication among clinicians and communication with clients or other laypersons. It is important to devise summaries of suicide-related nomenclatures that are of equal length, or else there is a risk that ease of usage will dominate clinicians’ ratings of perceived clinical utility. With these caveats in mind, this dissertation next proceeds to the hypotheses section.

**Hypotheses**

Having reviewed the literature regarding the two nomenclatures, reliability, and clinical utility, the next step is to formally present my hypotheses for this research. Overall, I hypothesize that the C-CASA will outperform the Silverman nomenclature, both in terms of its reliability and in terms of its scores on a measure of perceived clinical utility.

**Hypothesis 1:** The C-CASA is significantly shorter and simpler than the Silverman nomenclature; this will result in better reliability statistics since it will be easier for participants to consistently decide which category applies to a vignette. Also, the C-CASA is the only nomenclature of the two nomenclatures to have previously tested its reliability, which was in the “almost perfect” range (Landis & Koch, 1977). This hypothesis will be tested via kappa statistics. Specifically, I hypothesize that the C-CASA will display at least a 0.1 or better performance in terms of kappa statistics, in comparison to the Silverman nomenclature; in addition, the two nomenclatures’s kappas will display a statistically significant difference, as measured by 95% confidence intervals.

**Hypothesis 2:** Because the C-CASA is so much simpler in terms of the number of its categories than the other Silverman nomenclature, I hypothesize that it will score significantly better than the Silverman nomenclature on a measure of participants’ perception of its clinical utility. Specifically, I hypothesize that a one-way ANOVA will display significant omnibus differences between the two nomenclatures; C-CASA will have significantly higher mean scores on the perception of clinical utility measure than the Silverman nomenclature.
CHAPTER III

METHOD

Procedure

This dissertation employed a quantitative method to approach the problem of comparing the reliability and perceived clinical utility of two major current suicide-related nomenclatures. As described in prior chapters, this quantitative method was best suited to increase the reproducible nature of research concerning suicide-related nomenclatures; suicide-related nomenclatures have typically been studied using expert workgroups or workshops, both of which do not ordinarily permit replication of their methods. Specifically, this dissertation used electronically distributed surveys that contained one of the nomenclatures. Participants were asked to apply that nomenclature to vignettes that were written solely for the purpose of this research. Vignettes were designed so that only one term and definition would apply from each of the nomenclatures; all participants applied a nomenclature to the same set of vignettes. After participants applied the nomenclature to vignettes, they were asked to complete a perceived clinical utility survey designed to measure participants’ perception of how easy it was to use the nomenclature, how they believed the nomenclature would contribute to communication, and how they believed the nomenclature would contribute to treatment planning.

This procedure was used to collect data from two groups of participants; each group was presented with a different nomenclature to apply to the vignettes; however, all participants completed the same survey of perceived clinical utility. Each group had the same number of participants. Participants were randomly assigned to apply one of the two nomenclatures to vignettes.
After data was collected, the reliability of the nomenclatures was compared utilizing kappa statistics. Since prior studies have not assessed the psychometric properties of the perceived clinical utility questionnaire, this research assessed the measure’s structure via an exploratory factor analysis. The measure displayed a coherent structure, so the two groups’ mean scores on the measure were compared using a one-way ANOVA. The purpose of comparing groups’ mean scores was to see whether there are significant differences in the perception of clinical utility of each of the two nomenclatures.

Participants

Participants consisted of clinicians who are engaged at least five hours per week in the delivery of psychological services to clients. Participants were recruited primarily from the American Psychological Association’s divisional listservs, thus making psychologists the primary type of clinician surveyed. Specifically, an e-mail recruiting participants was sent to the following APA divisional listservs: Division 12 (Society of Clinical Psychology); Division 13 (Society of Consulting Psychology); Division 17 (Society of Counseling Psychology); Division 29 (Psychotherapy); and Division 42 (Psychologists in Independent Practice). The search for clinicians was also expanded to those on the suicidology listserv. Finally, clinicians were also sought via a snowball process through the use of the author’s friends on Facebook. However, the focus of clinician recruitment was those in clinical practice who are not suicidology experts (i.e., the “ordinary clinician”). Regardless of method of recruitment, in order to encourage participation all participants were entered in a drawing for one $200 gift certificate to Amazon.com.

A power analysis of a one-way ANOVA with two groups indicated that 128 participants were required to achieve a power of .80, assuming an alpha of .05 and a medium effect size (Cohen, 1988). For an exploratory factor analysis with 13 variables, the 10 participants per variable rule required 130 participants (Velicer & Fava, 1998); furthermore, 130 is more than the 100 minimum participants needed to conduct an exploratory factor analysis (MacCallum, Widaman, Zhang, & Hong, 1999). Therefore, at least 130 participants were recruited.
Demographic information was collected from the participants. The specific demographic questions asked of the participants are presented in Appendix C. They included gender, race/ethnicity, and marital status. In addition, participants were asked how many hours per week they were engaged in psychological service delivery. Psychological service delivery was defined to include direct client care; providing supervision to others regarding their direct client care; receiving supervision regarding direct client care; taking calls in a mental health call center or hotline; writing case notes regarding client care; conducting assessments for the purpose of direct client care; writing assessment reports for the purpose of direct client care; preparing case conference materials regarding client care; participating in case conferences regarding client care; and consultation with other providers regarding client care. In short, psychological service delivery included any activity done for the purpose of promoting direct client care; this excluded research and assessments that were not conducted for the purpose of promoting direct client care (such as academic research about psychological service delivery). If the participant did not provide at least five hours a week in direct client care, the participant was not be permitted to participate in the research. The purpose of this restriction was to ensure that participants are currently working clinicians, in order to prevent participants from answering perception of clinical utility questions from a merely theoretical point of view.

Participants were also asked to estimate approximately how many clients they have seen over the course of their career who have had suicide-related thoughts or behaviors. This served as an indicator of how experienced the participant was with working with suicide-related thoughts or behaviors. Finally, participants were asked what discipline they are licensed in, such as psychologist, marriage and family therapist, psychiatric nurse, psychiatrist, professional counselor, or social worker. Descriptive statistics regarding the makeup of the participant sample were reported, including number and percentages of participants across all demographic categories.
Overview of Statistical Procedures

This research concerned itself with the reliability and perceived clinical utility of the two current major suicide-related nomenclatures. To measure reliability, the kappa statistic was the most frequently used statistic in psychological and psychiatric research that concerns raters’ level of agreement with each other (Zanarini et al., 2000). To measure perceived clinical utility, an exploratory factor analysis was performed on an instrument created for this study; the instrument built on a measure which has no published psychometric statistics (Mullins-Sweatt & Widiger, 2011), thus making conducting an exploratory factor analysis advisable. Finally, assuming that the exploratory factor analysis revealed a stable, simple structure, the mean group scores of the two nomenclatures were compared with each other using a one-way ANOVA.

Measurement of Interrater Reliability

The kappa statistic was invented to gauge the reliability of two or more raters’ judgments: “The statistics kappa (Cohen, 1960) and weighted kappa (Cohen, 1968) were introduced to provide coefficients of agreement between two raters for nominal scales.” (Fleiss, Cohen, & Everitt, 1969, p. 323). The Silverman nomenclature and C-CASA are themselves nominal scales, in that each nomenclature’s categories is not ranked in a hierarchical fashion (as an ordinal scale would be), and they separate vignettes into independent categories (rather than numerically meaningful scales such as interval or ratio).

According to Fleiss, Cohen, and Everitt (1969), kappa was the correct statistic to use, rather than weighted kappa: “Kappa is appropriate when all disagreements may be considered equally serious, and weighted kappa is appropriate when the relative seriousness of the different possible disagreements can be specified.” (p. 323). Because we are interested in whether different raters would put the same material into the same nomenclature category, all disagreements between raters would be equally serious; that is, any disagreement between raters would be problematic in terms of the reliability of a particular
nomenclature. Therefore, kappa was the appropriate statistic to use to measure inter-rater reliability of each nomenclature.

The kappa statistic is regularly used for hypothesis testing in psychiatry, especially for calculating the likelihood that independent raters utilizing a diagnostic scheme have a larger percent agreement than would be predicted by chance alone; in fact, Zanarini et al. (2000, p. 292) described kappa values as “the standard measure of reliability in psychiatry because they correct for chance agreements.” The kappa statistic is defined mathematically as the observed probability of agreement \( (P_o) \) minus the expected probability of agreement \( (P_e) \), divided by 1 minus the expected probability of agreement:

\[
\kappa = \frac{P_o - P_e}{1 - P_e}
\]

Visually, the kappa statistic is most simply understood as a table in which one rater’s responses are represented by columns and the other rater’s responses are represented by rows. Observed agreement between the two raters can be found on the diagonal in such a table. A sample table is found below and will be used to explain the kappa statistic concretely.

For example, let us assume that each of two raters categorizes sixty observations into four categories. For each observation, there are sixteen possible outcomes, four of which involve rater agreement and twelve of which involve rater disagreement. In the sample table below, agreement is represented along the diagonal and is in boldface. Disagreement is represented by all other cells in the table: for example, there were eight observations in which Rater One placed each observation into category 2 while Rater Two placed each observation into category 1 and there were six observations in which Rater One placed each observation into category 1 while Rater Two placed each observation into category 2, and so forth for each of the twelve cells representing rater disagreement.

Using the sample table (Table 1) below, the kappa statistic can be readily understood. The observed probability of agreement is found by summing the diagonal: \( 2 + 3 + 4 + 7 = 16 \) cases in which both raters actually agreed in their categorizations. These 16 cases are divided by the total number of cases (60), for an observed probability of \( P_o = 0.27 \). The expected probability of agreement is computed by
multiplying the total number of observations in category 1 for Rater One by the total number of observations in category 1 for Rater Two, adding that to the total number of observations in category 2 for Rater One by the total number of observations in category 2 for Rater Two, and so forth, and dividing the sum by the square of the total number of observations: in the above example, the expected probability would be calculated as \( P_e = [(16\times16)+(17\times15)+(10\times14)+(17\times15)]/(60\times60) = 0.25. \)

Table 1

*Sample Table to Assist in Understanding Kappa*

<table>
<thead>
<tr>
<th></th>
<th>Category 1</th>
<th>Category 2</th>
<th>Category 3</th>
<th>Category 4</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Category 1</strong></td>
<td>2</td>
<td>8</td>
<td>4</td>
<td>2</td>
<td>16</td>
</tr>
<tr>
<td><strong>Rater Two</strong></td>
<td>6</td>
<td>3</td>
<td>1</td>
<td>5</td>
<td>15</td>
</tr>
<tr>
<td><strong>Category 3</strong></td>
<td>3</td>
<td>4</td>
<td>4</td>
<td>3</td>
<td>14</td>
</tr>
<tr>
<td><strong>Category 4</strong></td>
<td>5</td>
<td>2</td>
<td>1</td>
<td>7</td>
<td>15</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>16</td>
<td>17</td>
<td>10</td>
<td>17</td>
<td>60</td>
</tr>
</tbody>
</table>

Kappa in the above example can thus be readily computed as \( (P_o - P_e)/(1 - P_e) \), which is equal to 0.20 (using four decimal places for computation). According to Fleiss (1981), quoted in Zanarini et al. (2000) below, this value of kappa would represent poor agreement beyond chance:

For most purposes, values greater than .75 or so may be taken to represent excellent agreement beyond chance, values below .40 or so may be taken to represent poor agreement beyond chance, and values between .40 and .75 may be taken to represent fair to good agreement beyond chance. (p. 292)

An even more widely cited system for judging kappa values was given by Landis and Koch (1977), in which 0-0.20 represents slight agreement; 0.21-0.40 represents fair agreement; 0.41-0.60 represents moderate agreement; 0.61-0.80 represents substantial agreement; and 0.81-1.0 represents almost perfect agreement. The values chosen are ultimately arbitrary, so I used Landis and Koch’s system because it is
more widely accepted than Fleiss’s system. While the above example works for only two raters (i.e., Cohen’s kappa), Fleiss developed a kappa for more than two raters. Fleiss’s kappa does not require that the same raters categorize the observations – merely that the same number of raters categorize each observation. Since this provides more flexibility when there are many different clinician-participant raters (as here), Fleiss’ kappa will be used.

However, there are some well-known problems inherent in the kappa statistics described above: kappa is sensitive to prevalence and bias, each of which results in lowering kappa values. In this context, “a prevalence effect exists when the proportion of agreements on the positive classification differs from that of the negative classification.” (Sim & Wright, 2005, p. 261). Conversely, if a classification is either yes (present) or no (absent) (such as an SDVCS category would be yes or no, present or absent), there is no prevalence effect when there is a fifty-fifty chance that a rater will say that the category is present. When there is a prevalence effect, the probability that the category is present is either higher or lower than fifty percent. For instance, with the Silverman nomenclature, we would expect that any particular category would usually be absent rather than present (since there are 21 categories, only one of which would apply). Thus it is more likely, just by chance, that raters will agree when they both decide that a Silverman category is absent (i.e., they both decide no, a given category does not apply to a vignette), since the odds were already higher than fifty-fifty that both would choose no.

The prevalence effect can have a powerful influence on kappa. For example, Sim and Wright (2005, p. 261) gave an example in which the two raters agreed on the same number of cases but in which kappa was very different because of prevalence. In their examples, Sim and Wright created a scenario in which Clinician 1 and Clinician 2 agreed that a category was present 28 times and agreed that it was absent 2 times, while they disagreed 9 times. The kappa for this scenario, in which there was a strong prevalence effect (i.e., there was a 72 percent probability that the raters would agree just by chance), was .18 (poor). An alternative scenario was then examined, in which Clinician 1 and Clinician 2 agreed that a category was present 15 times and agreed that it was absent 15 times, while they again disagreed 9 times.
Notice that there is the same number of agreements (30) in both scenarios, but this time the kappa for the alternative scenario was .54 (moderate)! This is solely the result of chance agreement (which kappa adjusts for) being 50-50 rather than 72-28. Sim and Wright summarize this phenomenon: “When there is a large prevalence index, kappa is lower than when the prevalence index is low or zero.” (2005, p. 262).

Bias presents a similar problem as prevalence, but focuses on the pattern of disagreement between two raters rather than the pattern of agreement. Disagreements are asymmetrical (and bias is high) when the pattern of disagreement between two raters is not even. For instance, if there were 15 times in which Rater A chose “present” for a category when Rater B chose “absent,” while there were only 3 times in which Rater A chose “absent” when Rater B chose “present,” then disagreement is asymmetrical and bias is high. No bias would exist when the pattern of disagreement is symmetrical, e.g., 9 times in which Rater A chose “present” for a category when Rater B chose “absent,” and 9 times in which Rater A chose “absent” when Rater B chose “present.” Bias tends to inflate the value of kappa, while prevalence tends to deflate it. Whereas prevalence has the greatest effect on kappa when values of kappa are high, bias has the greatest effect on kappa when values of kappa are low (Sim & Wright, 2005).

When prevalence or bias is a problem, a “free-marginal” version of kappa can avoid the above problems. Randolph (2005) describes how free marginal kappa (or “$K_{\text{FREE}}$”) addresses the prevalence and bias issues:

One popular solution to the prevalence paradox in an agreement/reliability study with two raters is to assume that marginals are free and set $P_e$ equal to $1/k$, where $k$ is the number of rating categories. For example, if there are three rating categories, under the null hypothesis any two raters would be expected to agree on 1/3 of the cases. The free-marginal solution in the birater case has been suggested by many and has been shown to avoid the prevalence and bias paradox of high agreement but low kappa (Bennet et al. 1954; Brennan & Prediger, 1981; Byrt et al., 1954; Lawlis & Lu, and Maxwell, 1977). (p. 9)

For instance, the C-CASA has 8 categories, yielding $K_{\text{FREE}} = (P_o - 1/8)/(1 - 1/8)$. While $K_{\text{FREE}}$ avoids the prevalence and bias problems, Scott (1955) argues that when all categories do not have an equal probability of use, $K_{\text{FREE}}$ artificially inflates kappa values when there are a large number of categories. Randolph (2005), in contrast, believes that $K_{\text{FREE}}$ appropriately represents the probability of chance
assignment to a given category, in the same way that hypothesis testing compares actual results to results given by chance. Since there are good arguments for and against $K_{\text{FREE}}$ versus Fleiss’ kappa, therefore, the most conservative approach was to calculate both types of kappa and then to compare them.

I also reported descriptive statistics for the total number of vignettes classified, specifically the overall percent agreement among raters. Therefore, the statistics reported for kappa included the following: Fleiss’ kappa, $K_{\text{FREE}}$, and percent agreement.

**Hypothesis Testing and Power for Interrater Reliability**

In terms of hypothesis testing, the standard error of a sample of kappa distributions is known. Since the null hypothesis of $H_0 = 0$ would be extremely unlikely, it is preferred to set $H_0$ equal to a value (Walter, Eliasziw, & Donner, 1998). Since Landis and Koch (1977) consider 0.2 to represent slight agreement, I would argue that $H_0$ should be set equal to a value of 0.2. Furthermore, Landis and Koch (1977) consider 0.5 to represent the midpoint of moderate agreement, suggesting that $H_1$ should be set equal to a value of 0.5. Thus if a particular nomenclature demonstrated at least moderate agreement (0.41-0.60, Landis & Koch, 1977), the null hypothesis would be rejected. In fact, we would expect that actual inter-rater reliability would be at least 0.5 (i.e., $H_1 = 0.5$), given that the lowest reliability for categories of the C-CASA (the only nomenclature for which we have reliability data) is kappa = 0.47. With a null hypothesis of 0.2 and an alternative hypothesis of 0.5, a study with $N = 20$ or more (in fact, this study will have greater than 20 participants) requires 9.3 vignettes to achieve 80 percent power (Walter et al., Table II, p. 107). Rounding up for a slight cushion, this means that 10 vignettes was the optimal number in terms of power and efficiency. Those 10 vignettes were randomly selected from amongst the 29 available vignettes. This was done using www.random.org/integers/, which gave one a truly random set of numbers from among any specified set of numbers; the numbers 1-26 (corresponding to the numbering of the 26 vignettes) constituted the set and the program outputed 10 of those numbers, randomly. Those were the numbers of the vignettes that were used. This reduced the time burden to a more acceptable level for clinician-participants than if they were to rate all 26 vignettes.
Measurement of Perceived Clinical Utility

Mullins-Sweatt and Widiger (2011) utilized a six-question questionnaire to measure clinical utility. This same questionnaire was also used by Samuel and Widiger (2011) and Lowe and Widiger (2009) and was designed to measure aspects of clinical utility laid out by First et al. (2004). Unfortunately, no reliability or validity data exist for this six-question questionnaire. However, the six questions do address each of the three areas of clinical utility described in the literature review of the last chapter: ease of usage, communication, and treatment planning. In addition, the questions differentiate between communication between mental health professionals and communication with clients.

The six questions used by Mullins-Sweatt and Widiger (2011) were:

1. “How easy do you feel it was to apply the system to this individual?”
2. “How useful do you feel the system would be for communicating information about this individual with other mental health professionals?”
3. “How useful do you feel this system would be for communicating information about the individual to him or herself?”
4. “How useful is this system for comprehensively describing all the important personality problems the individual has?”
5. “How useful would this system be for helping you formulate an effective intervention for this individual?”
6. “How useful was this system for describing the individual’s global personality?” (p. 468)

Participants rated clients on a 5-point Likert-type scale in which 1 = Not at all useful; 2 = Slightly useful; 3 = Moderately useful; 4 = Very useful; and 5 = Extremely useful. However, this was difficult to adapt to the “ease of usage” questions; also, the use of seven-point Likert-type items is preferred to five-point items (Levant et al., 1992). Therefore, a seven-point Likert-type scale was used in which 1 = strongly disagree, 7 = strongly agree. This required changing the above questions into statements with which respondents would agree or disagree. Though one can see that the six questions were worded in order to assess DSM
and dimensional personality disorders, they were adapted to this dissertation’s vignette research. However, for purposes of this dissertation, the questionnaire should be referred to as a measure of *perceived* clinical utility, since clinician-participants were not asked to rate their actual experience with suicidal clients.

For the perceived clinical utility questionnaire, the six adapted questions were:

1. “It was easy to apply the nomenclature to these vignettes.”
2. “This nomenclature would be useful for communicating suicide-related information with other mental health professionals.”
3. “This nomenclature would be useful for communicating suicide-related information about a client to him or herself.”
4. “This nomenclature would comprehensively describe all the important suicide-related problems a client might have.”
5. “This nomenclature would be useful for helping you formulate an effective intervention for a client.”
6. “This nomenclature would be useful for describing a client’s global suicide-related issues.”

Because it would be excessively burdensome, in terms of time, to ask clinician-participants to answer questions after each vignette, the perceived clinical utility questionnaire was given once, at the end of all vignettes.

Other studies have used similar questions regarding clinical utility, though there have been a few variations. For instance, Rottman, Ahn, Sanislow, and Kim (2009) also used six questions, which were similar in terms of four of the six above questions; communication with other mental health professionals, communication with the client, comprehensive description, and global issues were similar. However, Rottman et al. (2009) also addressed two facets not addressed by Mullins-Sweatt and Widiger (2011): “How informative is this description in making a prognosis for this person?” and “How informative is this description in devising treatment plans for this person?” Translated into suicide-related nomenclature
language (and changing “informative” to “useful” to be consistent with the above), two additional questions were added:

7. “This nomenclature would be useful in making a prognosis for a client.”

8. “This nomenclature would be useful in devising treatment plans for a client.”

Whereas one would not normally add questions to an established measure, in fact there exists no reliability or validity information for any of the extant clinical utility measures. Since these questionnaires are therefore essentially at the exploratory phase of development, it seemed appropriate to pick and choose valuable aspects of each one to create a better clinical utility questionnaire.

Verheul (2005) added another useful question that ought to be addressed by a clinical utility questionnaire. Verheul (2005) noted the importance of assessing whether or not clinician-participants would actually use a particular system: “A classification system can be theoretically sound, valid, and provide perfect coverage, yet be completely worthless when it is not used at all or when it is not used correctly.” (p. 286). This implied that we ought to add a ninth question:

9. “Overall, it is likely that you would actually use this nomenclature in clinical practice.”

In terms of results, means and standard deviations were reported for each of the ten utility questions. For clinical utility, however, the primary purpose of this research was to test whether there were significant differences between each of the two nomenclatures. Therefore, a one-way analysis of variance (ANOVA) was used to compare mean clinical utility ratings that sum all ten questions of the perceived clinical utility questionnaire. This yielded a range of total perceived clinical utility scores of 9 to 63 summed points, with 10 representing the worst perceived clinical utility score (answering 1 to all questions) and 63 representing the best perceived clinical utility score (answering 7 to all questions).

**Exploratory Factor Analysis of Perceived Clinical Utility Measure**

Since no reliability or validity data exist for the proposed perceived clinical utility measure, it seemed appropriate to have some check on whether the perceived clinical utility measure in fact coheres the way it is intended. Exploratory factor analysis of the perceived clinical utility measure seemed well
suited to this purpose. However, in thinking about what factors might come out of the ten-item measure described above, it was readily seen that it was unlikely to yield a stable factor structure along the lines of ease of usage, communication, and treatment planning.

This can be seen when the nine questions for the proposed measure were put together in terms of their potential factors.

Factor One – Ease of Usage

“It was easy to apply the nomenclature to these vignettes.”

Factor Two – Communication

“This nomenclature would be useful for communicating suicide-related information with other mental health professionals.”

“This nomenclature would be useful for communicating suicide-related information about a client to him or herself.”

Factor Three – Treatment Planning

“This nomenclature would be useful for helping you formulate an effective intervention for a client.”

“This nomenclature would be useful in devising treatment plans for a client.”

“This nomenclature would be useful in making a prognosis for a client.”

Factor Four – Comprehensiveness

“This nomenclature would comprehensively describe all the important suicide-related problems a client might have.”

“This nomenclature would be useful for describing a client’s global suicide-related issues.”

“Overall, it is likely that you would actually use this nomenclature in clinical practice.”

However, the literature did not describe the importance of “comprehensiveness” as a potential fourth factor. Finally, two items on a factor did not provide stable structure (at least 3 items were required for
stable structure). Therefore, it appeared that the two items related to comprehensiveness (i.e., not the item about the overall likelihood of use of the nomenclature in clinical practice) should be eliminated.

However, since three items are required for a stable structure (per factor), it was prudent to include at least four items per factor (to allow for one item that does not work well). I therefore added 2-3 items to each of the three factors, taking care to create items that were similar in terms of the factor at issue, but not so similar as to be mere restatements of another question. I also tried to make the language of consistent between questions (relying on “How easy was it” for ease of usage questions and “How useful would this nomenclature be” for the remainder of the questions. The revised perceived clinical utility questionnaire is presented below, with its proposed factors in italics.

Perceived Clinical Utility Questionnaire

Ease of Usage:

1. “It was easy was to use this nomenclature.”
2. “It was easy was to apply this nomenclature to these vignettes.”
3. “It was easy to learn how to use this nomenclature.”
4. “It would be easy to teach someone else to use this nomenclature.”

Communication:

5. “This nomenclature would be useful for communicating suicide-related information to other mental health professionals.”
6. “This nomenclature would be useful for communicating suicide-related information to clients.”
7. “This nomenclature would be useful for communicating suicide-related information in writing.”
8. “This nomenclature would be useful for communicating suicide-related information by talking.”

Treatment Planning:
9. “This nomenclature would be useful for helping you formulate an intervention for a client.”

10. “This nomenclature would be useful in devising a treatment plan for a client.”

11. “This nomenclature would be useful in making a prognosis for a client.”

12. “This nomenclature would be useful in helping you conceptualize a client’s problems.”

**Overall Usefulness:**

13. “Overall, it is likely that this nomenclature would actually be useful in clinical practice.”

All questions were answered on a 7-point Likert-type scale in which 1 = Strongly agree and 7 = Strongly disagree. Therefore, a one-way analysis of variance (ANOVA) was used to compare mean clinical utility ratings that summed all thirteen questions of the perceived clinical utility questionnaire. This yielded a range of total perceived clinical utility scores of 13 to 91 summed points, with 13 representing the worst perceived clinical utility score (answering 1 to all questions) and 93 representing the best perceived clinical utility score (answering 7 to all questions).

If the perceived clinical utility measure were to demonstrate simple structure along the above lines, the thirteenth question could co-vary with any of the three major factors (perhaps thereby indicating that clinician-participants see that factor as the most likely to predict use of the nomenclature in clinical practice). Whichever factor the thirteenth item loads highest on, then that factor would have five items for calculating its alpha coefficient (as opposed to four items for the other two factors). If the perceived clinical utility questionnaire did not display good structure during the exploratory factor analysis, then descriptive statistics on single items would substitute for a one-way ANOVA. In other words, descriptive statistics of mean responses on each of the individual 13 questions would substitute for comparing group means via an ANOVA. Special attention would be paid to the six questions of Mullins-Sweatt and Widiger (2011), since others have relied upon those questions in prior research.

**Power Analysis for Perceived Clinical Utility**

A power analysis tells the researcher how many subjects are required to achieve a certain level of significance, given certain assumptions. In technical terms, power is equal to 1 – β, or one minus beta,
which is equal to Type II error. “Ideally, power should be at least .80 to detect a reasonable departure from the null hypothesis.” (Statsoft, 2012). One of the assumptions to calculate power is the effect size, or Cohen’s $f$. Cohen considered $f = .10$ to be a small effect, $f = .25$ to be a medium effect, and $f = .40$ to be a large effect. (Cohen, 1988). Another assumption is $\alpha$, which is typically .05 in psychological research (Wuensch, 2012).

Using G*Power v. 3.1.5 yields the following results for a one-way ANOVA with two groups. If the effect size is small (i.e., an $f = .10$), the total sample size required to achieve .80 power is 786 subjects. Obviously, this is unrealistic in terms of gathering participants. In contrast, if the effect size is medium (i.e., an $f = .25$), the total sample size required to achieve .80 power is 128. With two groups, this means 64 people in each group. The power analysis for a large effect size would yield a sample size of only 52 subjects. To be conservative, I assumed a medium effect size; I argue that based on the preeminence of the C-CASA in the research literature, as well as its dramatically less complicated structure (8 categories versus 21 categories for the Silverman nomenclature), expectation of at least a medium effect size is warranted. Therefore, I collected data from 128 clinician-participants (rounded up to 130).

**Open-Ended Exploratory Questions**

After presenting the perceived clinical utility questionnaire, the final aspect of this research was to ask some open-ended questions that might help guide future research. I proposed three simple questions to allow clinician-participants to contribute their overall impressions and thoughts about the nomenclature:

1. Please share the aspects of the nomenclature that you liked least, and why.
2. Please share the aspects of the nomenclature that you liked most, and why.
3. Finally, please share your overall thoughts about whether you would find this nomenclature to be helpful in clinical practice, and why.
Online Delivery Via SurveyMonkey.com

This vignette research lent itself to online delivery of its content via The University of Akron’s SurveyMonkey.com account. SurveyMonkey is an online questionnaire delivery system that permits the user to create different webpages and questions to be read on computer and answered by clicking on the appropriate answer choice. It is especially helpful for this type of research in that SurveyMonkey allowed the researcher to give participants a different web address for each of the two nomenclatures, while the content was identical in terms of vignettes. Also, since presentation of the nomenclature was tied to computer screen size, this allows roughly comparable lengths for each of the nomenclatures (i.e., one computer screen). Some scrolling was required to see all the answer choices for the Silverman nomenclature (due to the larger number of choices compared to the C-CASA), but in general the length of presentation was roughly comparable.

Summary of Method

This research was presented online to two groups of 64 clinician-participants each, for a total of 128 participants. Clinicians were recruited primarily from e-mails sent to the listservs of American Psychological Association divisions, from the American Association of Suicidology listserv, and by a snowball method from the author’s friends on Facebook; a $200 gift certificate to Amazon.com was given to one randomly selected participant to encourage participation in the study. The focus of clinician recruitment was those in clinical practice who are not suicidology experts (i.e., the “ordinary clinician”).

For those clinician-participants who agreed to participate, they were provided a web link to one of two websites via SurveyMonkey.com. Each clinician-participant provided informed consent by clicking “I agree” or “I do not agree” to proceed to the next screen after reading a summary of the risks and benefits of the research (and The University of Akron Institutional Review Board contact information). Demographic data was collected from participants, including gender, race, number of years in practice, psychological discipline, and whether they consider themselves to be an expert on suicide. In the main survey itself, the first screen presented the nomenclature that the clinician-participant applied to 10
vignettes, including background terms where relevant. It was explained that the gender of the people in
the vignettes was kept constant to eliminate variability due to perception of gender, as well as to make it
easy to identify the person at issue in the vignette (who is always female, whereas therapists or significant
others are always male). Next, the clinician-participant proceeded through 10 screens, one vignette per
screen, and was asked to choose one category in the nomenclature that best fit the information given in
the vignette. Next, the clinician-participant completed the clinical utility questionnaire, which contained
13 statements with which the clinician-participant agreed or disagreed with on a 7-point Likert-type scale.
Finally, the clinician-participant was asked to provide his or her open-ended impressions of the
nomenclature.

Across participants, Fleiss’ kappa, prevalence, bias, and free-marginal kappa were calculated and
compared. One-way ANOVAs of perceived clinical utility questionnaires allowed comparison of the two
nomenclatures, since the 10 vignettes were identical across nomenclatures (the 10 vignettes were
randomly selected from the 26 available vignettes, to reduce clinician-participant time burden). Means,
standard deviations, and percent agreement were also reported, where appropriate. Finally, themes were
reported for each of the three open-ended questions, thereby providing ideas for future research.
CHAPTER IV
RESULTS

Introduction

This dissertation concerned the comparison of two major current suicide-related nomenclatures, namely the C-CASA and the Silverman nomenclature. Having introduced the problem, summarized the literature, and reviewed the methodology, this chapter will present the results. This chapter was organized around the two hypotheses previously stated at the end of the literature review. Therefore, those hypotheses will be briefly restated below.

**Hypothesis 1:** The C-CASA is significantly shorter and simpler than the Silverman nomenclature; this will result in better reliability statistics since it will be easier for participants to consistently decide which category applies to a vignette. Specifically, it is hypothesized that the C-CASA will display at least a 0.1 or better performance in terms of kappa statistics, in comparison to the Silverman nomenclature.

**Hypothesis 2:** Because the C-CASA is so much simpler in terms of the number of its categories than the Silverman nomenclature, it will score significantly better than the Silverman nomenclature on a measure of participants’ perception of its clinical utility. Specifically, a one-way ANOVA will display significant omnibus differences between the two nomenclatures; C-CASA will have significantly higher mean scores on the perception of clinical utility measure than the Silverman nomenclature.

Since the first hypothesis concerns reliability statistics, those statistics will be presented first. This will be followed by an exploratory factor analysis of the perceived clinical utility questionnaire, to determine whether the factors of that measure are stable enough to rely upon in one-way ANOVA tests of differences between the two nomenclatures. Finally, the results of one-way ANOVAs will be presented.
Prior to those analyses, however, procedural and demographic characteristics of the sample will be described.

**Procedure and Demographics**

The ten random vignettes selected by the random integer generator at www.random.org were numbers 4, 5, 8, 9, 10, 17, 20, 22, 25, 26. Hence, the vignettes with those numbers were presented to participants. The order of presentation of measures was as follows: informed consent, demographics, vignettes, perceived clinical utility questionnaire, and open-ended questions. Each participant gave informed consent by typing his or her name at the end of the informed consent document. Prior to data collection, The University of Akron Institutional Review Board (IRB) certified that this research was considered exempt from review because the research represented minimal risk to participants. The IRB approval number was 20130112, and a copy of the IRB approval is presented in Appendix D.

A sample of “ordinary clinicians” was sought. Therefore, participants were recruited by e-mails to the listservs of three American Psychological Association divisions: Division 17 (Society of Counseling Psychology), Division 29 (Psychotherapy), and Division 42 (Psychological in Independent Practice). In addition, e-mail was sent to the listserv of the American Association of Suicidology. Finally, clinician-participants were solicited through the author’s Facebook page, in which he asked friends to pass on the solicitation message to clinicians that they knew. Overall, these methods combined to produce a final sample size of 131 clinician-participants.

However, initially there were 163 people who at least began the surveys. Therefore, 32 people began the survey but did not complete it, for a completion rate of 80.4%. Responses to the demographic questions and open-ended questions were not required by the survey software, while responses to the vignettes and perceived clinical utility questionnaire were required to advance through the survey. If a participant did not respond to a vignette or question that was required, he or she received an error message directing him or her to respond before the software would allow the participant to continue.
This resulted in complete data being obtained for vignettes and the perceived clinical utility questionnaire for those who finished the survey.

Data Cleaning

Before conducting any statistical analyses, the data were screened for missing values. Missing values in the demographic portion of the survey were ignored as inconsequential. Because of the procedure described immediately above (in which the software required answers to preceding questions before advancing to subsequent questions), it was straightforward to eliminate participants who did not completely fill out the survey with respect to vignettes and the perceived clinical utility questionnaire: the last question on the perceived clinical utility questionnaire was examined, and any participants who did not answer it were eliminated from the sample. This procedure resulted in 32 participants being eliminated from the sample, for a total sample size of 131.

The data were scanned for outliers by calculating z scores for each of the 13 continuous variables. Using Tabachnick’s and Fidell’s (2007, p. 73) cutoff of $z = 3.29$ (positive or negative) for outliers, there was only one outlier in the data. The one outlier’s value was changed from the value of “1” to the value of “2” so as to eliminate the extremeness of its score. Since the data from the ten vignettes represented (necessary) categories rather than continuous data, there could be no outliers among that data. As such, overall, no participants needed to be eliminated for outliers and only one data point was adjusted, rather than eliminated.

Demographics

The sample was 26.7% male and 73.3% female. Participants were predominantly Caucasian: 85.5% self-identified as white, 4.6% as Hispanic, 3.8% as Black, 3.1% as Asian, 1.6% as mixed race, 0.8% as American Indian, and 0.8% declined to answer. Participants were most often married: 61.1% were married, 20.6% were single (never married); 9.9% were single but in a long-term relationship, or engaged; 6.9% were divorced; 0.8% were separated or separating, and 0.8% declined to answer.
With respect to licensure, 74.8% indicated they were licensed in some professional discipline, while 25.2% answered “Other”; most of this latter category indicated that they were unlicensed students who were seeing clients as part of an internship, practicum, pre-doctoral or post-doctoral internship. In terms of mental health discipline, 42.7% were psychologists, 25.2% were unlicensed (mental health discipline was not clear since this category was listed as “Other” and most simply wrote “Student” as an explanation); 15.3% were professional counselors; 9.9% were social workers; 3.1% were marriage and family therapists, 2.3% were psychiatrists, and 1.5% were psychiatric nurses. Among the licensed participants, there were only a few who indicated that they were licensed in a different discipline than those listed above, such as drug and alcohol counseling or Master’s-level licensure. No one declined to answer this question.

With respect to hours worked in direct psychological service delivery each week, 17.6% worked 5-10 hours per week; 26.0% worked 11-20 hours per week; 24.4% worked 21-30 hours per week; 25.2% worked 31-40 hours per week; and 6.9% worked more than 40 hours per week. Everyone in the survey answered this question, so there were no missing values for it.

Participants were asked to estimate the number of clients they had worked with during their career who have had suicide-related thoughts or behaviors. There were 6.1% who declined to answer or who were unsure about an estimate. Of those participants who did provide an estimate, wide variability was seen in their answers, from a minimum of zero to a maximum of 5,000 clients. Breaking the responses down into rough groups, 12.2% estimated they saw 10 or fewer clients with suicide-related thoughts or behaviors; 37.4% estimated 11-99 clients; 6.9% estimated exactly 100 clients; 30.5% estimated 101-999 clients; and 6.9% estimated 1000 clients or more.

**Hypothesis 1: Reliability**

Reliability was calculated via kappa statistics, which measure the level of agreement between raters. It was hypothesized that the C-CASA nomenclature would demonstrate higher reliability than the Silverman nomenclature, specifically through statistical significance demonstrated by the non-overlap of
95% confidence intervals of the two nomenclatures, as well as by displaying a 0.1 higher kappa score than the Silverman nomenclature. This hypothesis was supported for both types of kappa, both fixed marginal kappa and free marginal kappa.

The C-CASA displayed an overall kappa of 0.58 (fixed marginal) and 0.59 (free marginal). In contrast, the Silverman nomenclature displayed an overall kappa of 0.42 (fixed marginal) and 0.45 (free marginal). This yielded a difference between the C-CASA and Silverman nomenclatures of 0.16 (free marginal) and 0.14 (free marginal). Furthermore, the 95% confidence interval for C-CASA’s fixed marginal kappa was 0.5728 to 0.5815 (given a calculated standard error of 0.00000492). This produced a Z = 260.11, which was significant at p<.001. Likewise, the 95% confidence interval for the Silverman nomenclature was 0.4216 to 0.4263 (given a calculated standard error of 0.00000145). This produced a Z = 351.55, which was significant at the p<.001 level. Since these 95% confidence intervals did not overlap between C-CASA and Silverman, there was a statistically significant difference between the C-CASA nomenclature’s kappa and the Silverman nomenclature’s kappa. As a result, the first hypothesis regarding reliability was supported. In terms of the system describing reliability of Landis and Koch (1977), all of the above kappa values are in the “moderate” range; moderate agreement includes kappa values from 0.41-0.60, which includes the 0.42-0.45 values of the Silverman nomenclature and the 0.58-0.59 values of the C-CASA nomenclature.

Percent agreement statistics were also calculated for both nomenclatures overall. For C-CASA, the overall agreement of raters was 64.6%. In contrast, the Silverman nomenclature displayed an overall agreement of raters of 45.1%. This produced a difference of 19.5%, in C-CASA’s favor. Another way to look at those percent agreement statistics is to notice that C-CASA’s percent agreement is 43.2% higher than the Silverman nomenclature’s percent agreement (i.e., 19.5% divided by 45.1% equals 43.2% higher).

Kappas and percent agreement statistics were also calculated for each of the 10 vignettes separately. For C-CASA, kappas for individual vignettes ranged from 0.13 to 0.96, while percent agreement ranged from 23.7% to 96.9%. For the Silverman nomenclature, kappas for individual vignettes
ranged from 0.10 to 0.85, while percent agreement ranged from 14.2% to 85.6%. Overall kappas, individual vignette kappas, and percent agreement statistics are displayed in Table 2.

One can see from Table 2 that C-CASA displayed higher kappas for seven of the ten vignettes, while the Silverman nomenclature displayed higher kappas for three of the ten vignettes. The largest disparity occurred in Vignette 8 (numbered #22 in Appendix B), in which C-CASA displayed a kappa of .96 and the Silverman nomenclature displayed a kappa of only .10. In contrast, kappas were nearly identical for Vignette 5 (numbered #10 in Appendix B), in which C-CASA had a kappa of .86 and the Silverman nomenclature had a kappa of .85. Both in terms of individual vignettes’ kappas and overall kappas, then, the first hypothesis was supported.

**Hypothesis 2: Perceived Clinical Utility**

Hypothesis 2 concerned the perceived clinical utility questionnaire. Specifically, it was hypothesized that one-way ANOVAs would show significant differences in group means of scores on the perceived clinical utility questionnaire, with the C-CASA demonstrating significantly higher perceived

| Overall Kappas, Individual Vignette Kappas, and Percent Agreement Statistics |
|---------------------------------|-----------------|-----------------|-----------------|
| C-CASA                          |                 | Silverman       |                 |
| Kappa                           | % Agreement     | Kappa           | % Agreement     |
| Overall                         | .59             | 64.6            | .45             | 45.1            |
| Vignette 1                      | .93             | 93.8            | .82             | 82.8            |
| Vignette 2                      | .25             | 34.4            | .57             | 59.1            |
| Vignette 3                      | .32             | 40.8            | .42             | 44.3            |
| Vignette 4                      | .13             | 23.7            | .28             | 31.7            |
| Vignette 5                      | .86             | 87.9            | .85             | 85.6            |
| Vignette 6                      | .96             | 96.9            | .79             | 80.1            |
Vignette 7   .60   65.2   .32   35.1
Vignette 8   .96   96.9   .10   14.2
Vignette 9   .46   53.1   .29   32.2
Vignette 10  .46   52.8   .08   12.3

Note. Vignettes 1-10 refer to numbered vignettes in Appendix B. Specifically, Vignette 1 = #4, Vignette 2 = #5, Vignette 3 = #8, Vignette 4 = #9, Vignette 5 = #10, Vignette 6 = #17, Vignette 7 = #20, Vignette 8 = #22, Vignette 9 = #25, and Vignette 10 = #26. All kappas are free marginal kappas. All values were calculated using the Online Kappa Calculator (Randolph, 2008), available at http://justusrandolph.net/kappa/

clinical utility than the Silverman nomenclature. However, given that the perceived clinical utility questionnaire is a previously untested instrument, it made sense to subject it to an exploratory factor analysis prior to using its factors (subscales) as variables to be compared using ANOVAs.

Exploratory Factor Analysis

It was hypothesized that an exploratory factor analysis (EFA) would produce a stable structure. A priori, three factors were hypothesized, but any stable structure that would permit use of one-way ANOVA tests (based on factors rather than individual items) would be acceptable. However, the factors would need to be theoretically interpretable, rather than simply an agglomeration of theoretically unrelated items.

Before conducting the EFA, the data were examined to ensure that factor analysis would be an appropriate methodological tool to apply to it. At a minimum, a sample size of 100 is needed for factor analysis (Gorsuch, 1983; Kline, 1979; MacCallum, Widaman, Zhang, & Hong, 1999). This sample size of 131 is adequate based on that criterion. Furthermore, Hatcher (1994) recommended that the number of subjects should be at least five times the number of variables to be analyzed. Since there are 13 variables to be analyzed in the perceived clinical utility questionnaire, this would produce a minimum of 65 subjects, which this sample exceeded. Others have suggested that the number of subjects should be at least 10 times the number of variables (Everitt, 1975; Nunnally, 1978; Velicer & Fava, 1998). This sample is
just over 10 times the number of variables (i.e., items in the perceived clinical utility questionnaire) since 10 times 13 is 130, and this sample contained 131 participants.

Another way of measuring sampling adequacy is the Kaiser-Meyer-Olkin value, which traditionally should be larger than .6 for the data to be factorable (Kaiser, 1974); here, the Kaiser-Meyer-Olkin value was .92. Furthermore, Bartlett’s Test of Sphericity (Bartlett, 1954) was statistically significant; this suggested that variables were correlated with each other, which would be necessary for factor analysis to be appropriate. This provided further evidence that factor analysis could appropriately be applied to this data.

In terms of the factor analytic technique that was applied to the data, Principal Axis Factoring (PAF) was chosen as opposed to Principal Components Analysis (PCA). PAF was preferable in this instance because it examines common variance in an effort to determine the underlying latent structure of the data. PCA, on the other hand, would not simply examine common variance but would also include unique and error variance; if PCA were used it would not be searching for latent structure but would instead simply reduce the data using all types of variance (Kahn, 2006).

Furthermore, it was important to use Principal Axis Factoring for these data because PAF does not assume multivariate normality (Fabrigar, Wegener, MacCallum, & Strahan, 1999). This data had negative skew due to participants’ generally positive evaluation of both nomenclatures’ perceived clinical utility. Techniques that would assume multivariate normality (such as Maximum Likelihood) would therefore be inappropriate.

Finally, the factors were rotated obliquely rather than orthogonally because one would expect that any factors discovered would be correlated with each other. Specifically, Promax rotation was chosen. In social scientific data, it is generally expected that factors would be correlated, thus making orthogonal rotation less preferable than oblique rotation (Preacher & MacCallum, 2003).

To determine the number of factors to extract, Kaiser’s criterion – namely, retaining factors for which eigenvalues are greater than one – and examination of the scree plot were first assessed. The scree
plot suggested that one factor would be the appropriate number of factors to extract, since the biggest drop-off occurred after factor 1, such that the scree began at factor two (see Figure 1 below). The eigenvalue-greater-than-one heuristic, on the other hand, suggested that two factors were the correct number to extract. The eigenvalue of the first factor was 8.05; the eigenvalue of the second factor was 1.31, and the eigenvalue of the third factor was 0.72.

However, since there is admittedly some subjectivity in examining the scree plot, a parallel analysis was also performed. Parallel analysis is a data-driven procedure that is considered the best practice in determining the number of factors to retain (Ledesma & Valero-Mora, 2007). One hundred permutations of a 13-variable, 131 participant structure were calculated by the Monte Carlo PCA for Parallel Analysis (Watkins, 2000) online. The data showed that the first random eigenvalue of 1.56 was lower than the first actual eigenvalue of 8.05. In contrast, the second random eigenvalue of 1.41 was higher than the second actual eigenvalue of 1.31. This movement from higher than the random eigenvalues to lower than the random eigenvalues suggests that a one-factor structure best represents the data (Levant et al., 2012).

![Scree Plot](image)

Figure 1. Scree plot for exploratory factor analysis of the perceived clinical utility questionnaire
The One-Factor EFA Solution

The one factor accounted for 61.9% of the variance. The minimum loading allowable was set to .40, which did not result in the removal of any items. No factors were removed for cross-loadings, since there was only one factor. The resulting 13-item, 1-factor scale with factor loadings and alpha is shown in Table 3 below. The factor was named "Perceived Clinical Utility" since its first item was the overall usefulness question; also, the 13 items represented the totality of the perceived clinical utility

Table 3

Perceived Clinical Utility Questionnaire Factor and Loadings from Principal Axis Factoring (Pattern Matrix) for the One-Factor Solution

<table>
<thead>
<tr>
<th>Factor (Percent Variance Accounted for; Alpha)</th>
<th>Item</th>
<th>Loading</th>
</tr>
</thead>
<tbody>
<tr>
<td>Factor 1: Perceived Clinical Utility (61.9%; α = .95)</td>
<td>Overall, it is likely that this nomenclature would actually be useful in clinical practice.</td>
<td>.87</td>
</tr>
<tr>
<td></td>
<td>This nomenclature would be useful for communicating suicide-related information to other mental health professionals.</td>
<td>.84</td>
</tr>
<tr>
<td></td>
<td>This nomenclature would be useful for communicating suicide-related information in writing.</td>
<td>.83</td>
</tr>
<tr>
<td></td>
<td>It would be easy to teach someone else to use this nomenclature.</td>
<td>.82</td>
</tr>
<tr>
<td></td>
<td>This nomenclature would be useful for communicating suicide-related information by talking.</td>
<td>.79</td>
</tr>
<tr>
<td></td>
<td>It was easy was to use this nomenclature.</td>
<td>.77</td>
</tr>
<tr>
<td></td>
<td>It was easy to learn how to use this nomenclature.</td>
<td>.77</td>
</tr>
<tr>
<td></td>
<td>It was easy was to apply this nomenclature to these vignettes.</td>
<td>.77</td>
</tr>
<tr>
<td></td>
<td>This nomenclature would be useful in devising a treatment plan for a client.</td>
<td>.76</td>
</tr>
<tr>
<td></td>
<td>This nomenclature would be useful in making a prognosis for a client.</td>
<td>.70</td>
</tr>
</tbody>
</table>
This nomenclature would be useful for communicating suicide-related information to clients. .68

This nomenclature would be useful for helping you formulate an intervention for a client. .68

This nomenclature would be useful in helping you conceptualize a client’s problems. .66

questionnaire, so it seemed sensible to keep that title for the one factor. It was interesting to note that three of the four “communication” items were near the top of the factor loadings, namely, “This nomenclature would be useful for communicating suicide-related information to other mental health professionals,” and “This nomenclature would be useful for communicating suicide-related information in writing,” which were the second- and third-highest loading items, respectively. In fact, the fifth-highest loading item also related to communication, namely, “This nomenclature would be useful for communicating suicide-related information by talking.” Furthermore, the fourth-, sixth-, seventh-, and eighth-highest loadings were the four items that originally constituted the “ease of usage” proposed factor. The lowest-loading four items all consisted of the items that originally constituted the “treatment planning” proposed factor. Therefore, the overall one-factor solution essentially kept the order of the hypothesized original three-factor solution: “communication” loaded the highest, then “ease of usage” loaded second-highest, and finally “treatment planning” loaded the lowest. In addition, the highest-loading factor was the “overall” item, which is what one would hope for a one-factor solution.

The internal consistency reliability of the one-factor solution was assessed using Cronbach’s alpha. The alpha for the total 13-item scale was .95, which fell into the range of “excellent” (Ponterotto & Ruckdeschel, 2007).

ANOVA Comparison of the Two Nomenclatures

Given that there were stable solutions from the exploratory factor analysis, the two nomenclatures were compared via ANOVAs on the 1-factor composite scores. For the one-factor solution,
the factor consisted of the mean of all 13 items in the perceived clinical utility questionnaire. The second hypothesis was that the C-CASA would score significantly higher on its mean perceived clinical utility score than would the Silverman nomenclature.

The second hypothesis was supported by the ANOVA data for the one-factor EFA solution (see Table 4). Specifically, the ANOVA produced an $F = 21.49$ that was significant at the $p < .001$ level. Sums of squares, degrees of freedom, and other ANOVA statistics are presented in Table 5 below. The mean score of the C-CASA on the perceived clinical utility questionnaire was 5.43, while the mean score of the Silverman nomenclature was 4.62. Higher scores represent greater perceived clinical utility on a 7-point Likert scale, where 4 represented “neutral” and 7 represented “strongly agree” to the statements that concerned positive perceived clinical utility. The Silverman nomenclature, then, received positive scores that were a bit about neutral, while the C-CASA received scores almost a point higher, which were almost halfway between “neutral” and “strongly agree.”

Table 4

ANOVA Statistics for the One-Factor Solution’s Mean Scores on Perceived Clinical Utility

<table>
<thead>
<tr>
<th></th>
<th>Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between Groups</td>
<td>21.31</td>
<td>1</td>
<td>21.31</td>
<td>21.49</td>
<td>$p &lt; .001$</td>
</tr>
<tr>
<td>Within Groups</td>
<td>127.94</td>
<td>129</td>
<td>0.99</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>149.25</td>
<td>130</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. df = degrees of freedom; Sig. = significance

Overall, therefore, hypothesis two was supported. Means of C-CASA on the perceived clinical utility scale were significantly higher than means of the Silverman nomenclature on the perceived clinical utility scale. Descriptive statistics summarizing these means and 95% confidence intervals are presented in Table 5 below.
Effect size was calculated based on the differences between the group means and the average of the standard deviations of the two groups. From Table 5, one can see that the mean for C-CASA for the one-factor solution was 5.43; the mean for the Silverman nomenclature was 4.62; and the standard deviations were 0.82 and 1.14, respectively. The average of the two standard deviations, then, was 0.98. The difference between the group means was 0.81, divided by the average standard deviation of 0.98 yielded an effect size of 0.83. Based on the rules of thumb for Cohen’s effect size, this is considered a large effect size since it is above 0.8 (Cohen, 1988). With a large effect size, two groups, a known sample size (131), and alpha = 0.05, the power to detect mean differences in this study was better than 0.99 (which would have required a sample size of 118, less than this study’s sample size of 131).

Table 5

*Descriptive Statistics for Scores on the Perceived Clinical Utility Questionnaire*

<table>
<thead>
<tr>
<th>One-Factor: Factor 1</th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
<th>Std. Error</th>
<th>95% CI</th>
<th># Items</th>
</tr>
</thead>
<tbody>
<tr>
<td>C-CASA</td>
<td>64</td>
<td>5.43</td>
<td>0.82</td>
<td>0.10</td>
<td>[5.22, 5.64]</td>
<td>13</td>
</tr>
</tbody>
</table>

*Note.* Std. Error = Standard Error; # Items = number of items included in the one-factor scale. Factor 1 of the one-factor solution is Perceived Clinical Utility.

**Open-Ended Questions**

Three open-ended questions were posed to clinician-participants: what did you like least about the nomenclature; what did you like most about the nomenclature; and would this nomenclature actually be useful in clinical practice. Themes and examples noted for each of the three open-ended questions are presented below.
The method of discerning patterns and overlap between responses was subjective. Qualitative coding with multiple raters was not attempted; rather, the themes noted are ideas for future research as opposed to a thorough characterization of all responses.

**C-CASA: Liked Least**

Forty-two participants gave responses for this category. For C-CASA and the “liked least” open-ended question, there was great variability in responses. The largest number of responses had to do with specific concerns that were not replicated in other responses. For example, one participant noted, “I do not believe that suicidal ideation should include ‘passive thoughts of death.’ I also think that ‘Not Enough Information’ leaves an ‘out’ for clinicians who either don’t know how to do a more thorough risk assessment or do not bother to do so.” Another participant stated, “I don’t like how interrupted and aborted suicide attempts were collapsed with other types of preparatory behavior.” A third participant said, “Some of it differs from more commonly used nomenclature. I might be concerned that if I were to use the nomenclature without further explanation, other professionals might misunderstand what I am trying to report.” A fourth participant wrote, “I think the nomenclature is helpful; however, it does not convey the amount of risk for suicide.”

However, there were some themes that were replicated between participants. For example, five participants said there was nothing they did not like about C-CASA. Another theme which five participants converged upon concerned “gray areas,” e.g., “With all nomenclatures there is a ‘gray’ space that people fit into. Some people felt like they might be in between terms.” Another participant wrote, “Didn’t allow for more qualitative (gray area) info that could clarify distinctions between categories.” Another theme which three participants endorsed was that the definitions were “too wordy” (e.g., “the items were a bit wordy”) or “too hard to remember” (e.g., “too complicated to remember the descriptions”). Finally, another theme that emerged was that the C-CASA oversimplified complex suicidal phenomena; four participants made statements consistent with this theme. For example, one participant wrote, “Could
potentially oversimplify some very complex suicidal situations,” while another stated, “Does not allow for nuanced clinical formulation,” and a third said, “Doesn’t cover the variability of suicidal behavior.”

**Silverman: Liked Least**

Fifty-six people gave responses in this category. For the Silverman nomenclature, there were three clear themes that emerged. In terms of the themes identified, two rose to the top: first, the idea that it was difficult to differentiate between some of the items, most often between suicidal communications and suicidal threats; second, that it was too complex; and third, that the last vignette was difficult to categorize with the Silverman nomenclature.

For the first theme, examples were: “I had a hard time distinguishing between suicidal threats and suicidal communications”; “It was difficult to distinguish between a suicidal threat, communication, and maybe one of the other options”; “The definitions of suicidal communications were difficult for me to understand. It seemed difficult to differentiate these from suicidal threats and suicidal ideation”; “The difference between suicidal threat and communications is unclear”; “The difference between suicide-related communication and threat needs more clarification.” There were other examples that were omitted for the sake of brevity; overall, ten responses fell into this theme.

For the second theme of “too complex/confusing,” examples were: “I think it would be too hard to remember all the types and distinctions”; “The number of statements could be overwhelming”; “It is too complicated. I found that there were too many descriptors and that “Type I” and “Type II” suggestions just confused me more”; “Possibly a few too many options; might take a [while] to really learn and memorize”; “somewhat confusing”; “There were too many options and it seemed that many of them were similar. It appeared to me that the number of terms could be reduced.” Overall, there were nine responses that captured the idea that the Silverman nomenclature was too complex/confusing, often noting that there were too many categories.

The third theme that emerged was that people had a difficult time categorizing the last vignette. Since this related to vignette construction, all of the responses that fell into this category (there were five)
are reproduced below. The final vignette was intended to be a suicide plan for the Silverman nomenclature and an accident for C-CASA (no one stated difficulty with this vignette for C-CASA). As a reminder, the text of the final vignette presented in the survey was:

A woman is leaving her therapist’s office after talking about her plan to kill herself next week. Distracted by her thoughts about what her therapist and she were discussing, she misses a step on the stairs and falls down a flight of stairs. She is taken to the hospital for lacerations to her legs and is given x-rays to see if anything is broken.

Statements concerning participants’ confusion about the vignette included: “Did not like the nomenclature for the final vignette – nothing appeared to communicate a potential unrelated accident...”; “For the last vignette, it would be unclear how to categorize the woman’s accident”; “I also would have liked a choice for none. For example, the woman who accidentally tripped down the stairs was not trying to kill herself although she did have that conversation earlier”; “On #10 [the last vignette] there seemed to be no good description of what took place”; “There was not an applicable option for the last vignette.” Though these five responses were negative about the last vignette, it should be noted that 62 participants did not comment on having any problems with the final vignette.

C-CASA: Liked Most

Forty-four participants gave responses in this category. One of the themes regarding what participants liked most about the C-CASA was that it was easy to use and understand: “Quick and easy to use”; “Provided easy groups”; “Pithy, good range of options”; “Quick, survey did not take long to complete”; “Concise descriptions”; “Easy to understand, made sense”; “It was easy to read, use, and apply”; “It’s good for shorthand psychotherapy notes.” Ten statements fell into this theme of ease of understanding and usage.

Another theme was that participants liked the way that the C-CASA tried to delineate different types of suicidal thoughts, behaviors, and self-injurious thoughts and behaviors. Representative statements included: “Differentiating between preparatory behavior versus ideation”; “Great effort in trying to clarify distinctions in suicidal/self-destructive thoughts and behaviors”; “I like that the nomenclature tries to separate out self-injury from suicide attempts...”; “I like the breakdown of the types
of thoughts”; “I liked the distinctions between intent to kill oneself or harm oneself and lack of intent”;
“Intent and action are separated”; “It distinguishes important features such as intent, action, thinking.”
Ten statements fell into this theme of distinction/delineation of different types of thoughts and behaviors.

A final theme included the clarity of its terms and definitions. Representative statements included: “Clear definitions based on intent to harm”; “Clarity of terms”; “Clarifies a murky subject”; “An attempt to clarify behaviors that have meanings along many dimensions”; “Having a clearly identified categorization could be helpful”; “Rather unambiguous”; “The clearly defined, differentiated labels for different levels of self-harm”; and “The nomenclature was descriptive which allows clinicians to be very clear.” These eight statements made up this third category of improving clarity in definitions and terms.

Silverman: Liked Most

Fifty-eight participants gave responses in this category. The first theme was that the different categories of the Silverman nomenclature improved specificity. For instance, “I appreciated the specificity of the nomenclature”; “I believe the variety of statements provided gave me a more specific way to categorize...”; “I liked the specificity of each description”; “It is very specific”; “It really delineated all the different possibilities”; “Very specific and precise”; “The precision”; “Very specific and provided a broader way to think about suicide...” These eight statements constituted the first theme of specificity in the Silverman nomenclature.

The second theme – and the dominant one in terms of numbers of participants – was that the Silverman nomenclature provided ways to differentiate degrees of categories such as injuries, intent, and plan from each other. Representative statements included: “Degrees of severity”; “Differentiating between with and without injuries”; “Differentiating the levels of intent and plan for the different categories”; “Having the types and distinguishers within each cluster is helpful”; “Degrees of each category”; “I do like the way that ideation, plan, and intent were made separable from whether or not there were injuries”; “I favored the manner in which some of them were broken down into categories of severity”; “I like that both outcome (injury, death, etc.) and intent were considered”; “Focus on intent
rather than outcome”; “I liked the breakdown of intent, thought, and action, in each grouping”; “I liked the different levels for each phase”; “Parsing out the difference between self-harm and suicide and the differentiation in levels of suicidality”; “Parsing things out with degree of intent is critical.” Overall, fifteen statements fell into this theme.

**Overall Helpfulness in Clinical Practice: C-CASA**

Forty-four participants responded to this question about the overall helpfulness of the C-CASA in clinical practice. The most basic difference was between those statements that were positive about versus negative about utilizing the C-CASA in clinical practice. Six statements were unambiguously negative; six statements were ambiguous (presented both positive and negative aspects, or were unclear); and the remainder (thirty-one) were positive.

People were generally more verbose with these answers than with the “liked least” and “liked most” open-ended answers, which makes it more difficult to list some representative answers. However, the thirty-one positive statements included things like improving a common language (six statements fell into this category, e.g., “[It] help[s], establishes a common language”; “I use this nomenclature because it is useful in communicating with other professionals”; “Yes, it gives a common language for clients, therapist[s] and others involved to use so it is clear what the risk is”), improving documentation (three statements fell into this category, e.g., “I believe this would be helpful… for documentation purposes”; “I think for documentation purposes, and for research purposes, it is helpful”; “Yes it would be [useful] in documenting suicidal ideation, plan, and attempt(s)...”), and overall summary statements (six statements fell into this category, e.g., “Absolutely”; “Anything is helpful with suicide – especially concrete things like this”; “At the early stages of practice this would be great for a therapist...”; “Great start in prompting thought along these distinctions!”; “Helpful in communicating to other providers and in discussing suicidality with clients”; “I would find it helpful to use with my clients and colleagues”).

In terms of the six ambivalent/ambiguous answers, the following are representative: “Although this nomenclature has the potential to be helpful in clinical practice, I would not want to see it replace the
narrative explaining personalized situations”; “I think it would be helpful in communicating with other professionals as a kind of shorthand. However, it can be overly simple if I want to discuss the case as a whole.”; “It isn’t really too different from what I do anyway, in a less formal way.” Other qualified responses were too long to be shared here.

In terms of the six negative responses, the following are representative: “I don’t know that these labels would be helpful to clients, and [I] also wonder if clinicians (particularly those not in the field of suicidology) could or would utilize these terms”; “I probably would not find it helpful, it is too complicated than just writing or saying my own brief explanation for a patient’s suicidal behavior”; and “[The] nomenclature is a bit complex in vocabulary and overload of answers that are provided.” One person stated that he or she prefers a different formalized nomenclature.

**Overall Helpfulness in Clinical Practice: Silverman**

Fifty-eight people responded to the question about the overall helpfulness of the Silverman nomenclature in clinical practice. There were 14 negative responses, 13 ambivalent/ambiguous responses, and 31 positive responses.

The largest theme involved in negative responses was a lack of simplicity, which included six of the 14 responses. For example, representative responses included: “A simple definition of suicide, ideation, plans, and attempt[s] with modifiers regarding intent and injury is much better”; “I feel that it is very difficult to use due to it being rather cumbersome”; and “If the nomenclature was pared down to the big categories (ideation, communication, plan, etc.) – otherwise there are too many names/categories.”

For the ambivalent/ambiguous responses, the largest theme was that participants gave at least one positive reason to use the nomenclature followed by at least one negative reason. This type of response included 9 of the 13 responses. Representative responses included: “Helpful for communication purposes, notes, and liability but generally not helpful for conceptualizing or treatment planning”; “I find it moderately helpful; good for conceptualization but a little impractical”; “I think for communication between clinicians it would be helpful, but I’m not so sure about communication with clients”; and “I
think it could be useful but it feels a bit bulky and I’m not sure medical professionals would use it faithfully.” As one can see, the content of the ambivalent statements varies, but the form of listing reasons pro and con was relatively consistent.

For positive responses, there were two themes. The first theme is that the Silverman nomenclature would help to clarify thoughts and behaviors in clinical settings. Representative statements included: “For conceptualization, yes, very much so”; “Defines the behavior and thoughts in finer detail”; “I think it would be helpful in providing a more concrete conceptualization of clients...”; “It can [help] pinpoint how the client is feeling”; “[It] helps to clarify [suicidal] intent.” Seven responses included this positive theme.

The second theme was that the Silverman nomenclature would be helpful in assessing suicide risk or dangerousness to self. Representative statements included: “It would make it easier to assess risk”; “Helpful... in terms of communication about the degree of concern [for a client]”; “In consulting about assessing risk”; “It is helpful for me to categorize the client’s suicidal risk”; “It would be good to have standardized ways of specifically communicating what the client has experienced... as well as the lethality involved”; “This nomenclature forces the clinician to consider some of the fundamental indices of overall suicidal risk...”; “It offers a detailed yet simplified way to think about the nuanced areas of suicide assessment and determination of lethality.” Ten statements included this positive theme.

Conclusion

This chapter presented the analysis of the data from 131 clinician-participants who completed the surveys. Both quantitative hypotheses were supported. First, the C-CASA’s reliability was indeed more than 0.1 higher for its kappa than was the Silverman nomenclature’s, and the difference between the two nomenclatures’ kappas was statistically significant. Second, the exploratory factor analysis of the perceived clinical utility questionnaire revealed a stable structure that permitted comparison of the two nomenclatures’ scores on the perceived clinical utility questionnaire. Third, there were significant differences between the two nomenclatures in terms of their scores on the perceived clinical utility
questionnaire. These differences were in the expected direction, with the C-CASA’s scores being significantly higher (more positive) than those of the Silverman nomenclature. Finally, themes from the responses to the three open-ended questions were presented; these themes may provide food for thought in the next chapter, which will discuss and interpret these results.
CHAPTER V

DISCUSSION

Introduction

This chapter will discuss and interpret the results presented in the prior chapter. The two major current suicide-related nomenclatures were empirically compared based on clinician-participants’ application of the nomenclatures to ten hypothetical vignettes. The nomenclatures were compared on two dimensions: reliability and perceived clinical utility. The two hypotheses of this dissertation received support on both dimensions: the C-CASA displayed greater reliability and better perceived clinical utility than did the Silverman nomenclature.

These results provide empirical evidence regarding terms and definitions that may be of use to clinicians and researchers in the field of suicidology. Since the modern advent of the study of suicide approximately a half-century ago, nomenclatures have been primarily judged and modified based on discussions in expert workgroups. While such workgroups have had their place in the development of promising nomenclatures, the field continues to be plagued by the absence of consensus in the language of suicidology (Maris, 2002). Therefore, it seemed to be time for major current nomenclatures to be subjected to empirical scrutiny via replicable survey research. In so doing, it was hoped that progress would be made in reaching consensus on a common language that could be used across clinical and research settings.

Part of the difficulty in empirically analyzing terms and definitions that make up a nomenclature is that it was tough to separate out the benefits of particular terms in the nomenclature. Thus, this research focused on comparing the two nomenclatures as a whole, as opposed to comparing their constituent terms and definitions. Finally, the three open-ended questions provided ideas for future
research and food for thought regarding potential causes behind the data collected. While these open-ended questions were not empirically analyzed in the way that reliability and perceived clinical utility were, they provided some insight into participants’ experience of applying the nomenclatures to the vignettes. As such, examination of the themes of the responses to the open-ended questions could provide useful directions for further research questions.

**Reliability of the Two Nomenclatures**

According to Landis and Koch (1977), both nomenclatures’ reliability scores were in the “moderate” range, namely between 0.4 and 0.6. However, the C-CASA’s reliability was near the top of this range, at 0.59, while the Silverman nomenclature was closer to the bottom of that range, at 0.45. C-CASA’s reliability was much lower than the 0.89 found in the original study by Posner et al. (2007). There could be at least two possible explanations for the current study’s lower reliability scores. First, the original study concerned narratives of actual adverse events in pharmaceutical trials; it could be that those actual events were less ambiguous than were the hypothetical vignettes presented in this study. Second, the original study had only nine raters, all of whom were expert suicidologists. It would make sense that nine experts would have greater consensus (on either actual adverse events or hypothetical vignettes) than would the raters in this study. The raters in this study were not expert suicidologists: in fact, their level of experience with suicidal clients ranged widely, from less than 10 clients to hundreds of clients with suicidal thoughts or behaviors seen during the clinicians’ careers. Furthermore, with that wide range of experience, there was in addition a much larger number of raters, which we would expect would lower reliability scores: specifically, there were 64 raters for C-CASA in this study, compared to nine raters in Posner et al. (2007).

In addition, the nine expert suicidologists in Posner et al. (2007) reviewed 423 adverse event cases, which is a much larger number of cases than the ten vignettes classified in this study. It could be that over many cases, reliability improved because the impact of particularly ambiguous cases would be attenuated. In this study, on the other hand, there were three vignettes that had percent agreement (for
C-CASA) of 23.7%, 34.4%, and 40.8%, respectively. This is in contrast to the other seven vignettes, in which percent agreement ranged from 52.8% up to 96.9%. Put in terms of kappa statistics, reliability ranged from 0.13-0.25 for the three vignettes, whereas it ranged from 0.46-0.96 for the other seven vignettes. If those three vignettes represented particularly ambiguous cases (unlike narratives about real adverse events), this could lower the overall kappa scores for the C-CASA in this study.

However, it could also be that the 423 adverse event narratives that the nine expert suicidologists rated were very unambiguously written. This could be a function of drug companies wanting to keep their adverse event reporting as brief as possible, so as to not emphasize the flaws in their medications. This theory is speculative, however, especially since we do not have access to the 423 adverse event narratives rated in the Posner et al. (2007) article. That is one advantage of having created ten vignettes rather than relying upon 423 adverse events: others who might read this dissertation would have access to all of the cases rated since the vignettes are provided here.

However, while we cannot straightforwardly compare the C-CASA’s data here to that of Posner et al. (2007), we can compare C-CASA’s data here to the Silverman nomenclature’s data in this study. Whereas C-CASA’s reliability was much lower here than it was in Posner et al. (2007), it was undoubtedly higher here than was the Silverman nomenclature’s. Specifically, C-CASA’s higher kappa of 0.59 compared to the Silverman nomenclature’s 0.45 translated into percent agreement statistics that are perhaps more intuitively understood: C-CASA had 64.6 percent agreement, whereas the Silverman nomenclature had only 45.1 percent agreement. In other words, C-CASA had nearly two-thirds agreement, while the Silverman nomenclature had less than half agreement. This appears to be a substantial difference.

In fact, this difference becomes even more apparent when kappas and percent agreements for the ten individual vignettes are examined. It was mentioned above that seven vignettes (for C-CASA) had kappas of .46 or higher, with percent agreements ranging from 53.1% to 96.9%. In contrast, only four vignettes (for the Silverman nomenclature) had percent agreements of greater than 50%. That means that a full six vignettes (for the Silverman nomenclature) had percent agreements less than 50%: two vignettes
were less than 15% (14.2% and 12.3%); three vignettes were in the thirties (31.7%, 32.2%, and 35.1%); and one vignette was in the forties (44.3%). Therefore, the variability of reliabilities was seemingly large and relatively poor, compared to C-CASA’s. Comparing the high end of kappas and percent agreement did not help the Silverman nomenclature either. Whereas for C-CASA, three vignettes had kappas and percent agreements higher than .93 and 93.8%, respectively, the Silverman nomenclature did not have any vignettes that were in the nineties, either by kappa or percent agreement. Therefore, on both the high end and the low end of kappas and percent agreements for individual vignettes, the Silverman nomenclature underperformed the C-CASA.

In summary, then, the C-CASA outperformed the Silverman nomenclature in terms of reliability statistics. On the individual vignette level, this indicated that raters agreed more often about which category a particular vignette should be classified into. On the overall level, this indicated that raters agreed more often overall about which categories all vignettes should be classified into. This suggested that, for these vignettes at least, ordinary clinicians showed greater agreement using the C-CASA than they did using the Silverman nomenclature.

The meaning of this greater reliability varies based on clinical versus research settings. In research settings, reliability of a nomenclature is critical to relying upon that nomenclature for research conclusions. If participants could not even agree how to use a nomenclature in the same way, it would suggest that the nomenclature might not be measuring the same thing between participants. In other words, validity presupposes reliability; without reliability as a basic premise, it would be difficult to conclude that the nomenclature measures what it claims to be measuring, since people could not even agree on how to apply the nomenclature.

In clinical settings, agreement has less to do with validity than it has to do with clinical utility. Without reliability between clinicians, it would be difficult for clinicians to have a useful dialogue using the nomenclature; they do not even agree about how to use the nomenclature. Using the terms and definitions of the Silverman nomenclature, for instance, would give rise to agreement less than half the
time between clinicians. This assumes that this vignette research would generalize to actual clinical use, of course, but it is an illustrative point nonetheless. While the C-CASA would have its own problems in terms of agreement among clinicians, it would at least approach two-thirds agreement most of the time.

Assuming that these results would generalize to clinical and research settings, these reliability statistics have highlighted the potential insufficiency of our two major current suicide-related nomenclatures. While the C-CASA performed better in terms of reliability statistics, it still only performed in the moderate range (Landis & Koch, 1977). In practical terms, it would result in agreement less than two-thirds of the time. This would suggest that even the C-CASA could use improvement in order to make the nomenclature more reliable.

The kinds of changes that might improve the C-CASA’s reliability are unclear, however. Open-ended responses about what participants liked least, for example, about the C-CASA included that the definitions were “too wordy,” suggesting that sparser definitions might assist people in applying the nomenclature more easily. However, one of the themes that people liked most about the C-CASA was that it was easy to use and understand, including that descriptions were “concise” and “pithy.” These results were buttressed by those on the perceived clinical utility questionnaire, in which participants were halfway between “neutral” and “strongly agree” on items that included ease of usage. Furthermore, an examination of the definitions themselves in C-CASA revealed that the definitions ranged from one to three sentences in length. It would be difficult to pare those definitions down further without losing clarity. In fact, it could be argued that longer definitions would increase reliability, but then the perceived clinical utility would probably decrease, since participants liked the brief nature of the definitions.

Finally, as mentioned above, it could be that the difficulty in reliability has to do with the nature of the vignettes rather than any insufficiency in the terms and definitions themselves. Confronted with real-life situations, and greater details about the situations than the brief vignettes provided, participants might have been able to apply the nomenclature more reliably. Increasing the length and detail included in vignettes might be one way to improve reliability, thereby making them closer to real life situations in
which clinicians have a wealth of data about clients. In fact, one might develop vignettes in concert with clinicians, asking them about real-life cases that seemed to exemplify each of the categories in C-CASA. Furthermore, to improve consistency among vignettes, one could conduct focus groups in a pilot project intended to tease out those areas of vignettes that were unclear and thereby contributed to a lack of reliability in responding.

There is one last possibility that should be mentioned, however. It is possible that moderate agreement is all that is realistic when a range of ordinary clinicians utilize any nomenclature. There could be too much variability among clinicians as people to permit them to agree on classifying vignette – or real-life situations – into the inherently gray areas of language concerning suicide-related phenomena. The fact that real life cases contained gray areas was one of the themes in the open-ended questions. One could test this hypothesis by allowing only one type of clinician to participate in the survey, perhaps with only one level of experience. For example, one might specify that participants must be licensed psychologists with between five and ten years of experience seeing clients. Any such restriction, though, makes it more difficult to obtain participants in real-life research.

**Perceived Clinical Utility**

Perceived clinical utility was measured using a new instrument that was theoretically based. Theoretically, this instrument had three factors: communication, ease of usage, and treatment planning. However, exploratory factor analysis revealed that the actual instrument in fact seemed to have one factor, which consisted of all thirteen of the items in the original questionnaire. In terms of factor loadings, the highest loading item was the item, “Overall, it is likely that this nomenclature would actually be useful in clinical practice.” This is what one would want from a one-factor scale measuring perceived clinical utility: the overall clinical utility item was the highest loading item. Furthermore, the rough structure of the three theorized factors was preserved in that the original hypothesized communication subscale’s items were the highest loading items for the one-factor model; the loadings of the ease of usage items were in the middle range, and the treatment planning items loaded near the bottom. In total,
loadings were strong, with no loadings less than 0.65, and the top loading was above 0.85. The alpha for
the scale was .95, which is obviously in the excellent range (Ponterotto & Ruckdeschel, 2007), suggesting
that the items all measure the same construct of perceived clinical utility.

With this one-factor questionnaire, ANOVAs revealed significant differences between the two
nomenclatures at the p < .001 level. On the perceived clinical utility questionnaire, higher scores
represented greater perceived clinical utility on a 7-point Likert scale, where 4 represented “neutral” and
7 represented “strongly agree” to the statements that concerned positive perceived clinical utility. The
Silverman nomenclature received positive scores that were a bit above neutral (4.62), while the C-CASA
received scores almost a point higher (5.43), which were almost halfway between “neutral” (4) and
“strongly agree” (7). Another way of looking at this spacing between neutral and strongly agree might be
that participants agreed that the C-CASA would have perceived clinical utility, while they were about
neutral on whether the Silverman nomenclature would have perceived clinical utility.

In terms of effect size, Cohen’s d was above 0.8, and therefore is typically interpreted as a large
effect size (Cohen, 1988). This was consistent with the practical difference between a result slightly above
“neutral” (the Silverman nomenclature) and one that was about halfway between “neutral” and “strongly
agree” (C-CASA). In real world terms, this might mean that the C-CASA would be perceived as helpful to
clinical practice, while the Silverman nomenclature would be neither helpful nor unhelpful. This is indeed
an important result, consistent with the large effect size.

This would suggest that the C-CASA might be seen as useful for everyday clinicians who would
apply this language to their daily clinical concerns. Based on the items in the perceived clinical utility
questionnaire, this would imply that clinicians would find the C-CASA helpful in terms of overall usefulness
in clinical practice, in communication with clients and other mental health professionals, and in terms of
ease of usage of the nomenclature. In other words, the C-CASA was perceived as helpful across a range of
clinical utility issues. In contrast, the Silverman nomenclature was seen as close to neither helpful nor
unhelpful for this same range of clinical utility concerns.
The upshot of this difference between the C-CASA and the Silverman nomenclature might be that the C-CASA would be more promising in working towards a standardized common language for suicidology among ordinary clinicians. While the clinician-participants in this sample did not “strongly agree” that the C-CASA would be useful for clinical practice, they did roughly agree that it would be helpful. Since the purpose of measuring perceived clinical utility was to judge which nomenclature might make a better contribution to the development of a common language for clinicians and researchers, this would suggest that the C-CASA would be the instrument with more promise in that regard.

Directions for Future Research

Given that the C-CASA performed better in terms of both reliability and perceived clinical utility, this would suggest that future research could be focused upon the C-CASA. However, prior to doing so it would behoove us to replicate the results of this study. One option for doing so would be to utilize the vignettes that were not randomly selected – but which were created – for this research. In other words, this study could be done again, but this time using ten new vignettes that were also created for this research. In the alternative, this research could be exactly duplicated, but with a new sample and by a new researcher. Replication would ensure greater confidence in these results.

Assuming that the results would replicate well, this would suggest that the Silverman nomenclature does not work as well in real-world clinical practice than does C-CASA. To be more accurate, one might have said that clinicians perceive this. However, given the need for a nomenclature that clinicians would generally use, the perception of clinical utility would translate straightforwardly into the reality of clinical utility. This would be because a nomenclature that clinicians would not use would be useless in terms of creating a common language for clinicians.

Other directions for future research would include replicating this research with researchers rather than clinicians. While clinicians would constitute the larger number of professionals concerned with suicide-related phenomena, researchers would add incrementally to our knowledge base concerning suicide and self-harm. If researchers would not use C-CASA, this would hamper acceptance of a
nomenclature that might gain universal acceptance. In doing that research, of course, items in the
perceived clinical utility questionnaire would need to be changed from “clinical practice” to “research” to
allow for appropriate translation of meaning for researchers.

Another line of research would concern the perceived clinical utility questionnaire itself. Now
that an exploratory factor analysis has been completed, the factor structure of the new instrument could
be further tested in several regards. First, since larger sample sizes are generally helpful for EFA, the
factor structure of this new instrument could be re-tested with more participants; this would also allow
for replication of the instrument’s factor structure. Second, a confirmatory factor analysis could be
conducted, perhaps with competing one- and two-factor structures for the perceived clinical utility
questionnaire.

If confirmatory factor analysis were to support the one-factor structure of this new instrument,
research could be conducted on whether the instrument would generalize to other clinical contexts. For
example, the word “nomenclature” could be changed to “clinical theory,” “strategy,” “documentation
method,” “supervision method,” or any other clinical measurement that could have perceived clinical
utility. Representative examples would include, “This clinical theory would be useful in devising a
treatment plan for a client”; “This strategy would be useful for communicating information to clients”; or
“It was easy to learn how to use this supervision method.” Some creativity would be required in order to
translate each of the items into another format, but it might be a productive endeavor to see whether the
instrument could be helpful with other clinical issues besides suicide and nomenclatures.

In fact, the perceived clinical utility questionnaire might be adapted to measure the perceived
clinical utility of assessment measures. For example, if one were to develop a new measure that
measured a new construct relevant to clinical practice, one might want to get an idea of whether this new
measure would hold promise in actual practice. If the measure displayed an absence of perceived clinical
utility (measured by scores of “neutral” down to “strongly disagree”) this could save time and effort in
validating the new instrument, since it would be rational to abandon an instrument that clinicians would
not likely adopt. On the other hand, extra research effort might be devoted to developing instruments that clinicians would perceive as helpful to practice.

Returning to a discussion of future research directions for suicide-related nomenclatures, another tack that could be productive would be a formal qualitative study of the aspects that participants found helpful and unhelpful in the C-CASA. Here, the open-ended questions elicited a few themes, but those themes were not detailed enough to allow for much application to the C-CASA itself. Furthermore, qualitative examination of focus groups might allow for specific suggestions in revising particular terms and definitions. Those revisions could focus on increasing reliability and agreement among clinicians in application of the nomenclature, or they might focus on improving the perceived clinical utility of the terms and definitions of the C-CASA.

There were some hints about improving the perceived clinical utility of the C-CASA given by the themes in the open-ended questions. One important strength of the C-CASA, according to participants’ responses to the “liked most” section of the open-ended questions, is that it was easy to use and understand. Specifically, participants appreciated that the C-CASA was “pithy” and that terms and definitions were “concise.” Therefore, one of the important aspects of the C-CASA was that definitions read simply and were not too long. This suggested that improvements to the C-CASA should not involve making the instrument more wordy or longer; instead, if new categories were added or existing categories were changed, it would be important to retain the C-CASA’s simple and easy-to-use current structure. Furthermore, the “liked least” section of the open-ended responses to the Silverman nomenclature reinforced this theme: participants did not like the plethora of terms and definitions offered by the 21-category Silverman nomenclature. Therefore, care should be taken in revising the C-CASA to ensure that the number of categories would not balloon into an excessively large number of categories. Rather, attention should be paid to improving existing categories, rather than adding a number of new categories.
A related line of research might approach additions and subtractions to the C-CASA by analyzing clinical case notes to see what terms clinicians actually use most in clinical practice. Speculation in this regard might include the addition of a category of “suicide plan.” If this category were added, however, care would need to be taken to ensure that its definition would be easy to understand and concise, given that those were attributes that participants especially appreciated about the C-CASA in the “liked most” section of the open-ended questions. Adding a category, therefore, would be a tricky endeavor that should be handled with care. In addition, this method would have the difficulty of obtaining access to clinical case notes because of privacy concerns. Therefore, clinicians might be surveyed regarding the terms that they use most in clinical practice. The theory in this regard might be that it is easier to induce standardized language among clinicians by mirroring the language they currently use, rather than attempting to get them to adopt a new nomenclature such as the C-CASA. Alternatively, one might discover that clinicians already use terms that are similar to those of the C-CASA, thereby improving support for its likely adoption by clinicians.

An upshot of the qualitative results was that there was apparently a delicate balance between specificity of a nomenclature and its ease in use and understanding. In other words, participants appreciated both the concise nature of the C-CASA and the specific distinctions of the Silverman nomenclature. However, participants also liked the differentiations of different categories within the C-CASA (e.g., the specificity of looking at suicidal “intent”). Therefore, future qualitative exploration might attempt to tease apart the areas in which the C-CASA could use more specificity, without sacrificing too much of its concise, easy-to-understand terms and definitions.

Limitations of This Research

This research had a number of important limitations. Obvious limitations included the demographics of the sample, which was primarily White (85.5%), female (73.3%), and married (61.1%). Furthermore, there were more psychologists (42.7%) than any other type of clinicians; the next closest category was students (25.2%), followed by professional counselors (15.3%). Future efforts could be made
to recruit larger numbers of clinician participants from other disciplines, including professional counselors, social workers, marriage and family therapists, psychiatric nurses, and psychiatrists. Future research could endeavor to recruit more minority and male participants so as to obtain a more representative sample of the population. In addition, recruiting more minority clinicians would perhaps increase the likelihood that minority clients would be represented; though African-Americans have lower rates of suicide than do Caucasians, suicide was still the third-leading cause of death among African-American youth (McIntosh, 2013). Increasing minority clinician representation might be fostered by recruiting from Division 45 of the American Psychological Association, which is the Society for the Psychology Study of Ethnic Minority Issues. Also, about one quarter of participants were unlicensed students. Future research could attempt to recruit only licensed professionals, thereby ensuring that inexperience would not color the results; this could increase the likelihood that participants would have significant real world clinical experience to base their perceived clinical utility judgments upon.

Other substantive limitations have been remarked upon previously. The most important limitation was that of the hypothetical vignettes. First, by employing hypothetical vignettes, the generalizability to real world clinical examples was reduced. Second, these hypothetical vignettes were all created by this author, and were not pilot tested. Future research could employ several researchers utilizing a consensual process for creating vignettes, and then could pilot test those vignettes to improve them.

Another limitation was the use of a previously untested questionnaire to measure perceived clinical utility. This limits the confidence we could have in our results since the validity and reliability of the measure was unknown. Future research that would provide empirical support for that measure could improve the confidence we would have in our results, across the board.

An additional limitation was the lack of measurement of the reliability of particular categories within the C-CASA. This measurement was not attempted because it would presuppose that vignettes were sufficiently uniform and stable in their responses from participants that results would generalize to
the individual C-CASA categories rather than representing “noise” from vignette instability. Furthermore, as an initial study of the overall reliability and clinical utility of two competing nomenclatures, that level of detail seemed inappropriate. However, that could be a promising avenue for future research.

Another limitation was that the research was survey-oriented and measured clinicians’ perceptions of clinical utility rather than independent, non-subjective measures of that clinical utility. For example, examination of clinical case notes or video of clinical case conference would be direct approaches to measuring clinical utility, rather than simply measuring hypothetical self-report. However, given privacy concerns and the time- and effort-intensity of those methods, only survey data were collected here.

Finally, this research was the first of its kind in attempting to measure the reliability and perceived clinical utility of two competing suicide-related nomenclatures. This created several limitations in the research that were inherent to new research paradigms. First and foremost, this research did not have the benefit of prior research methods or assessment measures to build upon. Second, there were no applicable measures for perceived clinical utility, so an entire questionnaire had to be created, albeit based on prior theory; however, that theory was previously untested and lacked validity or reliability data. Third, this research was not able to build upon prior efforts to evaluate nomenclatures in a replicable way, since such prior efforts did not exist. Instead, nomenclatures have been previously advanced through expert workgroups whose processes were hidden and whose conclusions thus could not be replicated.
CONCLUSION

Suicide is one of the leading causes of death worldwide. The measurement of suicide-related phenomena has been hamstrung by the lack of an accepted, standardized nomenclature. Unlike clinical practice generally, in which the DSM is the dominant clinical nomenclature, during the past half-century there have been no widely utilized terms and definitions that most clinicians and researchers have agreed upon. This has created problems in measuring suicide (including epidemiology), in treating suicide and self-harm, and in researching suicide-related thoughts and behaviors.

Two of the major current suicide-related nomenclatures were the C-CASA and the Silverman nomenclature. While the C-CASA had previously had some data collected regarding its reliability, the Silverman nomenclature had not. Neither nomenclature had any data collected concerning perceived clinical utility. However, the C-CASA had been used in FDA analysis of adverse events for pharmaceutical products. Since the C-CASA had, therefore, some real world reliability and utility, it was hypothesized that vignette research would show greater support for the C-CASA’s reliability and perceived clinical utility than would the Silverman nomenclature. These hypotheses were supported for both reliability and perceived clinical utility; participants indicated that C-CASA would likely be helpful in clinical practice while the Silverman nomenclature was judged close to neutral in regards to its perceived clinical utility.

However, significant questions for future research remain. Would vignette research generalize to real world examples? Could this research be replicated by others? Can a link be made between the C-CASA nomenclature and predicting real-world variables such as attempting suicide? Ultimately, nomenclature would be most significant if it were to contribute to the classification and prediction of behavioral variables such as attempted suicide. These questions are significant and may spawn more research into the reliability and perceived clinical utility of suicide-related nomenclatures, particularly the C-CASA. Future explorations might focus upon improving the C-CASA in order to make it even more attractive to clinicians and researchers who need a common language to discuss, treat, and research suicide-related phenomena. The hope would be that advancement of our knowledge base with respect to
the language of suicidology would incrementally advance our ability to clearly and effectively combat one of the leading causes of death.
REFERENCES


American Psychological Association (2012). APA Practice Organization information alert: New psychotherapy billing codes for 2013: Fundamental services underlying new codes will not
change. Listserv of Division 17 of the American Psychological Association. E-mail received September 17, 2012, available upon request.


117


Kipnes, M. S., Hollander, P., Fujioka, K., Gantz, I., Seck, T., Erondu, N., . . . Amatruda, M. (2010). A one-year study to assess the safety and efficacy of the CB1R inverse agonist taranabant in overweight and


Thomas Reuters (2012). *Social Science Citation Index.* Posner et al. (2007) article’s citation map accessed August 16, 2012; Silverman et al. (2007a,b) article’s citation map accessed August 18, 2012; Brenner et al. (2011) article’s citation map accessed August 20, 2012.


126


APPENDIX A

PRESENTING THE TWO MAJOR CURRENT SUICIDE-RELATED NOMENCLATURES

The terms and definitions of the two major current suicide-related nomenclatures are presented below. As in the literature review, the Columbia Classification Algorithm of Suicide Assessment (C-CASA) will be presented first, followed by the Silverman nomenclature. In each case, terms and definitions are quoted from each of the relevant articles, but for ease of reading quotation marks will be omitted.

The Columbia Classification Algorithm of Suicide Assessment

Posner et al. (2007) presented the C-CASA in their article in the *American Journal of Psychiatry* entitled, “Columbia Classification Algorithm of Suicide Assessment (C-CASA): Classification of suicidal events in the FDA’s pediatric suicidal risk analysis of antidepressants.” In that article, they presented the C-CASA itself on pages 1037-1038. The following material encompasses quotations from those pages.

There were originally three columns in that article: “suicidal events,” “definition,” and “training examples.” The first two columns will be retained below, though “suicidal events” has been changed to “term” to better reflect the terms and definitions that make up a nomenclature, and “training examples” have been omitted for the sake of brevity.

Background Terminology

No background terminology is presented in C-CASA.

Terms to Be Classified

Term: Completed Suicide

Definition of Completed Suicide: A self-injurious behavior that resulted in fatality and was associated with at least some intent to die as a result of the act.

Term: Suicide Attempt
Definition of Suicide Attempt: A potentially self-injurious behavior, associated with at least some intent to die, as a result of the act. Evidence that the individual intended to kill him/herself, at least to some degree, can be explicit or inferred from the behavior or circumstance. A suicide attempt may or may not result in actual injury.

Term: Preparatory Acts Toward Imminent Suicidal Behavior

Definition of Preparatory Acts Toward Imminent Suicidal Behavior: The individual takes steps to injure him- or herself, but is stopped by self or others from starting the self-injurious act before the potential for harm has begun.

Term: Suicidal Ideation

Definition of Suicidal Ideation: Passive thoughts about wanting to be dead or active thoughts about killing oneself, not accompanied by preparatory behavior.

Term: Self-Injurious Behavior, No Suicidal Intent

Definition of Self-Injurious Behavior, No Suicidal Intent: Self-injurious behavior associated with no intent to die. The behavior is intended purely for other reasons, either to relieve distress (often referred to as “self-mutilation,” e.g., superficial cuts or scratches, hitting/banging, or burns) or to effect change in others or the environment.

Term: Other, No Deliberate Self-Harm

Definition of Other, No Deliberate Self-Harm: No evidence of any suicidality or deliberate self-injurious behavior associated with the event. The event is characterized as an accidental injury, psychiatric or behavioral symptoms only, or medical symptoms or procedure only.

Term: Self-injurious behavior, suicidal intent unknown

Definition of Self-injurious behavior, suicidal intent unknown: Self-injurious behavior where associated intent to die is unknown and cannot be inferred. The injury or potential for injury is clear, but why the individual engaged in the behavior is unclear.

Term: Not Enough Information
Definition of Not Enough Information: Insufficient information to determine whether the event involved deliberate suicidal behavior or ideation. There is reason to suspect the possibility of suicidality but not enough to be confident that the event was not something other, such as an accident or psychiatric symptom. An injury sustained on a place on the body consistent with deliberate self-harm or suicidal behavior (e.g., wrists), without any information as to how the injury was received, would warrant placement in this category.

The Silverman Nomenclature

Silverman et al. (2007a,b) presented the Silverman nomenclature in their articles in the journal Suicide and Life-Threatening Behavior entitled, “Rebuilding the tower of Babel: A revised nomenclature for the study of suicide and suicidal behaviors. Part 1: Background, rationale, and methodology;” and “Rebuilding the tower of Babel: A revised nomenclature for the study of suicide and suicidal behaviors. Part 2: Suicide-related ideations, communications, and behaviors.” In those articles, they presented the Silverman nomenclature itself in the main body text. Personal communication with Morton Silverman (August, 2012) helped clarify the terms and definitions further. The following material encompasses quotations from those articles.

Background Terminology

Term: Suicidal Intent

Definition of Suicidal Intent: Intent refers to the aim, purpose, or goal of the behavior, so suicidal intent refers to suicide as the aim, purpose, or goal of the person’s behavior. Although intent implies an action, the action itself is not a given (e.g., “I intend to kill myself by hanging” does not mean that the action of suicide by hanging has occurred).

Term: Suicide-Related Behavior

Definition of Suicide-Related Behavior: A self-inflicted, potentially injurious behavior for which there is evidence (either explicit or implicit) either that (a) the person wished to use the appearance of intending
to kill himself/herself in order to attain some other end, or (b) the person intended at some undetermined or some known degree to kill himself/herself.

**Terms to Be Classified**

**Term: Suicide**

Definition of Suicide: A self-inflicted death with evidence (either explicit or implicit) of intent to die.

Terms: Self-Harm, Type I (without injuries); Self-Harm, Type II (with injuries); Self-Inflicted Unintentional Death.

Definition of Self-Harm, Type I, Type II, and Self-Inflicted Unintentional Death: A self-inflicted, potentially injurious behavior for which there is evidence (either explicit or implicit) that the person did not intend to kill himself/herself (i.e., had no intent to die). Type I is without injuries; Type II is with injuries; Self-Inflicted Unintentional Death is with fatal outcome.

Terms: Suicide Attempt, Type I (without injuries); Suicide Attempt, Type II (with injuries)

Definition of Suicide Attempt, Type I and Type II: A self-inflicted, potentially injurious behavior with a nonfatal outcome for which there is evidence (either explicit or implicit) of intent to die. Type I is without injuries; Type II is with injuries.

Terms: Undetermined Suicide-Related Behavior, Type I (without injuries); Undetermined Suicide-Related Behavior, Type II (with injuries); Self-Inflicted Death with Undetermined Intent (fatal outcome)

Definition of Undetermined Suicide-Related Behavior, Type I, Type II and Self-Inflicted Death with Undetermined Intent: A self-inflicted, potentially injurious behavior where intent is unknown. Type I is without injuries; Type II is with injuries; Self-Inflicted Death with Undetermined Intent is with fatal outcome.

Terms: Suicide-Related Ideations, Type I (with no suicidal intent); Suicide-Related Ideations, Type II (with undetermined degree of suicidal intent); Suicide-Related Ideations, Type III (with some suicidal intent)
Definition of Suicide-Related Ideations, Type I, Type II, and Type III: Any self-reported thoughts of engaging in suicide-related behavior. Type I is with no suicidal intent; Type II is with undetermined degree of suicidal intent; Type III is with some suicidal intent.

Terms: Suicide-Related Communications, Type I (with no suicidal intent); Suicide-Related Communications, Type II (with undetermined degree of suicidal intent); Suicide-Related Communications, Type III (with some suicidal intent)

Definition of Suicide-Related Communications, Type I, Type II, and Type III: Any interpersonal act of imparting, conveying, or transmitting thoughts, wishes, desires, or intent for which there is evidence (either explicit or implicit) that the act of communication is not itself a self-inflicted behavior or self-injurious. Type I is with no suicidal intent; Type II is with undetermined degree of suicidal intent; Type III is with some suicidal intent.

Terms: Suicide Threat, Type I (with no suicidal intent); Suicide Threat, Type II (with undetermined degree of suicidal intent); Suicide Threat, Type III (with some suicidal intent)

Definition of Suicide Threat, Type I, Type II, and Type III: Any interpersonal action, verbal or nonverbal, without a direct self-injurious component, that a reasonable person would interpret as communicating or suggesting that suicidal behavior might occur in the near future. Type I is with no suicidal intent; Type II is with undetermined degree of suicidal intent; Type III is with some suicidal intent.

Terms: Suicide Plan, Type I (with no suicidal intent); Suicide Plan, Type II (with undetermined degree of suicidal intent); Suicide Plan, Type III (with some suicidal intent)

Definition of Suicide Plan, Type I, Type II, and Type III: A proposed method of carrying out a design that will lead to a potentially self-injurious outcome; a systematic formulation of a program of action that has the potential for resulting in self-injury. Type I is with no suicidal intent; Type II is with undetermined degree of suicidal intent; Type III is with some suicidal intent.
APPENDIX B

VIGNETTES

Below are the types of vignettes that are needed to completely represent each of the terms in both nomenclatures. Before each vignette, in italics, is a description of the types of evidence contained within each vignette. After each type of vignette is listed the term that would apply to it from each of the two nomenclatures, with the applicable nomenclature in parentheses after each term.

Vignettes and their summary statements were rationally derived by examining the types of terms in each nomenclature. Care was taken to ensure that within each nomenclature, the vignettes fell into only one of the categories, as opposed to two categories or more. This entailed a thorough understanding of each nomenclature and the types of terms contained within each.

For example, in order to decide which term applies to a particular vignette, both nomenclatures require that one decide whether or not “suicidal intent” is present, absent, or undetermined by the evidence in the vignette. Likewise, both nomenclatures require that one decide the outcome of behaviors (e.g., whether or not there were injuries present, or whether there was fatal outcome). On the other hand, only the Silverman nomenclature distinguishes between “communications” and behaviors without communication; however, in order that the Silverman nomenclature can adequately classify the vignettes, the presence or absence of communications are noted in the italicized summary statements below.

When the reader examines the vignettes below, he or she will notice that each vignette is described in terms of shared aspects and unique aspects of each of the two nomenclatures. These shared and unique aspects were created by a content analysis of the overlapping and unique aspects of the two nomenclatures. Specifically, terms and definitions in each of the two nomenclatures are described in terms of the content that is unique to a particular nomenclature, as well as in terms of the content that is
shared by the other nomenclature. These shared and unique aspects are summarized as letters A. through J. below, thereby representing ten categories of either shared or unique aspects of the nomenclatures.

These types of shared and unique aspects of the nomenclatures are listed immediately below:

A. Whether suicidal intent is present, absent, or unclear (shared by both nomenclatures).

B. Whether there are only thoughts present (e.g., suicidal ideation) (shared by both nomenclatures).

C. When there are only thoughts present, whether there are communications of those suicide-related thoughts to others (unique to the Silverman nomenclature, representing three terms within it).

D. Whether there are suicide-related behaviors present (shared by both nomenclatures).

E. Whether those behaviors result in injuries, no injuries, or fatal outcome (shared by both nomenclatures).

F. Whether the behavior is accidental (unique to C-CASA, representing one term within it).

G. Whether the behavior is only preparatory (unique to C-CASA, representing one term within it).

H. Whether there is a plan (unique to the Silverman nomenclature, representing three terms within it).

I. Whether there are suicidal threats (unique to the Silverman nomenclature, representing three terms within it).

J. Whether there is not enough information to decide whether the behavior is suicide-related (unique to C-CASA, representing one term within it).

The purpose of constructing these categories A. through J. was to allow the construction of vignettes in an exhaustive but efficient manner. In other words, the 26 vignettes exhaustively represent every category in both nomenclatures; they do so by being constructed to match each of the categories A. through J. above. However, the vignettes were efficiently constructed because the unique categories were combined with shared categories within a single vignette.
For example, the concepts of suicidal communications, suicide threats, and suicide plan (which were each unique to the Silverman nomenclature) each represented three terms within the Silverman nomenclature. C-CASA had three unique terms too: preparatory behavior, accident, and not enough information. On the other hand, both nomenclatures contain some version of the term “suicidal ideation,” the key aspect of which is that there are suicidal thoughts, as opposed to suicidal behaviors. Likewise, both nomenclatures contained the concept of suicidal intent, as well as whether the suicidal intent was present, absent, or unclear whether it existed in a particular instance.

**Formal Description of the Mechanics of Creating the 26 Vignettes**

In terms of the mechanics of creating the vignettes, the following methodology was used. Because the Silverman nomenclature has the largest number of categories, vignettes were created according to its terms and definitions first. Specifically, I began at the top of the list of terms and definitions of the Silverman nomenclature and worked my way down through the list, in order. Thus, the first vignette was “suicide” and the second vignette was “self-harm, Type I (without injuries),” “other,” since these were the first and the second term from the top, respectively.

Since the C-CASA has only eight categories, in contrast to the Silverman nomenclature’s 21 categories, there would inevitably be categories in the C-CASA that are represented more than once. It seemed unfair for participants to have more than one vignette to categorize into the same term for the C-CASA, while having exactly one vignette to categorize for each of the Silverman nomenclature’s terms. Therefore, five extra vignettes were created so as to make sure that the Silverman nomenclature also had multiple vignettes for the same terms within it. These extra vignettes increased the total number of vignettes from 21 (the number of terms within the Silverman nomenclature) to 26.

The terms with two vignettes created within the Silverman nomenclature were: Self-Harm, Type II (with injuries); Suicide Attempt, Type II (with injuries); Undetermined Suicide-Related Behavior, Type II (with injuries); Suicidal Ideation, Type II (with undetermined degree of suicidal intent); and Suicide Plan, Type II (with undetermined degree of suicidal intent). While the number (5) of extra vignettes was
ultimately arbitrary, care was taken to ensure that the terms that had two vignettes created for them were different kinds of terms (i.e., not two “suicide attempt” terms, or two “suicide plan” terms).

Finally, the eight terms in the C-CASA were incorporated into the existing 26 vignettes. This required some rewriting of vignettes so that the three unique categories of the C-CASA (namely, “not enough information,” “other, no deliberate self-harm,” and “preparatory acts toward imminent suicidal behavior”) could be represented within those 26 vignettes. For example, one vignette was re-written so that its content continued to represent “suicide plan” within the Silverman nomenclature, but represented “preparatory acts toward imminent suicidal behavior” in the C-CASA. This was possible since C-CASA did not have a term encompassing suicide plans, while the Silverman nomenclature did not have a term encompassing preparatory behavior. Since there were three types of suicide plans, the (unique to C-CASA) category of “accident” was also rewritten so that it constituted a suicide plan in the Silverman nomenclature, i.e., efficiency was achieved by not having to write an additional vignette for either the Silverman nomenclature or the C-CASA. Finally, the “not enough information” category of the C-CASA was combined with “suicidal ideation, type II (with undetermined degree of suicidal intent)” in the Silverman nomenclature. In this way, only 21 vignettes were required to represent all of the categories in both nomenclatures.

An additional five vignettes were created, however, so that the Silverman nomenclature could potentially duplicate categories when participants responded to vignettes. This was done instead of each vignette having a unique and different response, which would have been different than C-CASA; C-CASA necessarily had repeated categories since there were ten vignettes and only eight categories in C-CASA. In this way, procedural fairness was achieved.

After having created vignettes for each of the C-CASA and Silverman nomenclature terms, I then used A. through J. to categorize those vignettes. Having categorized a particular vignette, I then used the A. through J. categories to decide which terms and definitions fit each of those vignettes for both the Silverman nomenclature and the C-CASA. To ensure that a particular vignette accurately fit the definition
of the term that was indicated by the categorization using A. through J., I then re-read the detailed term and its definition within C-CASA and the Silverman nomenclature, noting whether the vignette’s features met the definitional elements of a particular term. If there was any vagueness when I re-read the vignette and the terms/definitions that applied to it, I reworked the details of the vignette slightly to make it more clear which term and definition applied from each of the two nomenclatures.

The 26 Vignettes Presented

Below are the 26 vignettes, numbered 1 through 26. Each is initially described by language that was crafted to be as neutral as possible with respect to the two nomenclatures (e.g., for vignette number 1, “death by suicide” is neutral between “completed suicide,” (C-CASA) and “suicide,” (Silverman). Next, in italics are the presence or absence (or in the case of suicidal intent, “unclear evidence of suicidal intent”) of categories A. through J.; finally, the terms that apply from each of the two nomenclatures are given, with parentheticals indicating which nomenclature goes with which term.

1. Death by suicide (includes A: suicidal intent is present; B: not only thoughts are present; C: no communications to others; D: suicide-related behaviors are present; E: behaviors result in fatal outcome; F: behavior is not an accident; G: behavior is not only preparatory; H: without a plan; I: without threats; J: sufficient information): Completed Suicide (C-CASA); Suicide (Silverman)

   A man discovers his wife shot to death with a gun in her hand. Nearby is a note, in her handwriting, that explains that she couldn’t go on living anymore because of the intense emotional pain she had been in.

2. Non-suicidal self-harm (A: suicidal intent is absent; B: not only thoughts; C: not only thoughts, so communications irrelevant; D: suicide-related behaviors are present; E: injuries are absent; F: behavior is not an accident; G: behavior is not only preparatory; H: without a plan; I: without threats; J: sufficient information): Self-Injurious Behavior, No Suicidal Intent (C-CASA); Self-Harm, Type I (without injuries) (Silverman)
A woman has difficulty feeling alive. However, she does not want to die, nor does she want to cause herself pain. She takes a knife and presses its edge up against her skin but does not break the surface of the skin. She has done this regularly for many years and as always, she doesn’t injure herself but rather feels a little bit more alive because of the potential for injuring herself. She puts the knife away, feeling more energetic.

3. **Non-suicidal self-harm (A: no evidence of suicidal intent; B: not only thoughts present; C: not only thoughts, so communications irrelevant; D: suicide-related behaviors are present; E: injuries are present; F: behavior is not an accident; G: behavior is not only preparatory; H: without a plan; I: without threats; J: sufficient information):** Self-Injurious Behavior, No Suicidal Intent (C-CASA); Self-Harm, Type II (with injuries) (Silverman)

   A woman scratches herself on her legs with a paperclip, drawing blood. Her intention is to feel something that will make her feel alive. After she sees the blood – which she had not intended to cause – she abruptly stops scratching herself and bandages the cut. She meant to scratch herself more, like she usually does, but she stopped because of the blood. She did not intend to kill herself, just to scratch herself.

4. **Non-suicidal self-harm (A: suicidal intent absent; B: not only thoughts; C: not only thoughts, so communications irrelevant; D: suicide-related behaviors are present; E: injuries are present; F: behavior is not an accident; G: behavior is not only preparatory; H: without a plan; I: without threats; J: sufficient information):** Self-Injurious Behavior, No Suicidal Intent (C-CASA); Self-Harm, Type II (with injuries) (Silverman)

   A woman asks a man out on a date and he politely refuses. She feels unattractive and feels like a failure, so she goes into her bathroom and takes out her razor. She makes several cuts on her abdomen until all she feels is the physical pain and the emotional pain is dulled. She did not intend to kill herself, just to cut herself.
5. **Non-suicidal self-harm, fatal** (A: suicidal intent is absent; B: not only thoughts; C: not only thoughts, so communications irrelevant; D: suicide-related behaviors are present; E: behaviors result in fatal outcome; F: not accidental; G: behavior is not only preparatory; H: without a plan; I: without threats; J: sufficient information): Self-Injurious Behavior, No Suicidal Intent (C-CASA); Self-Inflicted Unintentional Death (Silverman)

   A woman has a history of banging her head against the wall when she feels anxious and sad. She has never tried to kill herself and in fact just the night before told her boyfriend that she would never kill herself because she is too afraid of death. This morning she felt anxious and sad and began banging her head against the wall. Unfortunately, this time a blood vessel burst in her brain and she died.

6. **Suicide attempt** (includes A: suicidal intent is present; B: not only thoughts are present; C: not only thoughts, so communications irrelevant; D: suicide-related behaviors are present; E: injuries are absent; F: not accidental; G: behavior is not only preparatory; H: without a plan; I: without threats; J: sufficient information): Suicide Attempt (C-CASA); Suicide Attempt, Type I (without injuries) (Silverman)

   A man comes home in the middle of the day (when his wife expected him to be at work) and discovers his wife with a knife poised to cut her wrists. As soon as she sees him she begins sobbing and puts down the knife without cutting herself. Nearby he sees a suicide note she wrote to him explaining that she had gambled away their savings and had decided to kill herself.

7. **Suicide attempt** (includes A: suicidal intent is present; B: not only thoughts are present; C: no communications to others; D: suicide-related behaviors are present; E: injuries are present; F: interruption by self or other is present; G: behavior is not only preparatory; H: without a plan; I: without threats; J: sufficient information): Suicide Attempt (C-CASA); Suicide Attempt, Type II (with injuries) (Silverman)
A woman who was recently diagnosed with a terminal illness decides that she does not want to be in physical pain for months before her death. She puts a noose around her neck, intending to kill herself. She steps off a chair but as soon as she falls, she changes her mind and manages to free herself. However, she suffers bruising and scrapes on her neck.

8. Suicide attempt (includes A: suicidal intent is present; B: not only thoughts are present; C: no communications to others; D: suicide-related behaviors are present; E: injuries are present; F: behavior is not accidental; G: behavior is not only preparatory; H: without a plan; I: without threats; J: sufficient information): Suicide Attempt (C-CASA); Suicide Attempt, Type II (with injuries) (Silverman)

After experiencing a divorce, an alcoholic woman decides to end it all by jumping off her balcony. She is not drunk at the time and clearly remembers wanting to die. However, she survives the fall and ends up in the hospital due to broken legs.

9. Undetermined self-inflicted potential for injury (includes A: suicidal intent is unclear; B: not only thoughts are present; C: not only thoughts, so communications irrelevant; D: suicide-related behaviors are present; E: absence of physical injury; F: not accidental; G: behavior is not only preparatory; H: without a plan; I: without threats; J: sufficient information): Self-Injurious Behavior, Suicidal Intent Unknown (C-CASA); Undetermined Suicide-Related Behavior, Type I (without injuries) (Silverman)

A woman does cocaine. While high, she puts one bullet in his gun, spins the chamber, puts the gun to her head and pulls the trigger. It does not fire and she is unhurt. When asked by her therapist later whether she intended to kill herself, the woman states that she was high and doesn’t know what she was thinking.

10. Undetermined self-inflicted injury (includes A: unclear evidence of suicidal intent; B: not only thoughts present; C: not only thoughts, so communications irrelevant; D: suicide-related
behaviors are present; E: presence of physical injury; F: not accidental; G: behavior is not only preparatory; H: without a plan; I: without threats; J: sufficient information): Self-Injurious Behavior, Suicidal Intent Unknown (C-CASA); Undetermined Suicide-Related Behavior, Type II (with injuries) (Silverman)

When very inebriated, a woman pours alcohol on herself and sets herself on fire. As the pain becomes intense, she manages to put the fire out and call 911. She wakes up in the hospital with no recollection of whether she meant to kill herself or just to burn herself.

She only remembers feeling bad about herself, getting drunk, and putting out the fire.

11. Undetermined self-inflicted injury (A: includes unclear evidence of suicidal intent; B: not only thoughts present; C: not only thoughts, so communications irrelevant; D: suicide-related behaviors are present; E: presence of physical injury, F: behaviors are not accidental; G: behavior is not only preparatory; H: without a plan; I: without threats; J: sufficient information): Self-Injurious Behavior, Suicidal Intent Unknown (C-CASA); Undetermined Suicide-Related Behavior, Type II (with injuries) (Silverman)

A woman has a history of banging her head against the wall when she feels upset. She is fired from her job and hits her head against the concrete wall at the top of the stairwell at work. She knocks herself unconscious, where a coworker finds her and calls 911.

When she wakes up in the hospital, she refuses to answer any questions about whether she was trying to injure herself or kill herself. She just stares into space and refuses to talk.

12. Undetermined self-inflicted death (includes A: unclear evidence of suicidal intent; B: not only thoughts present; C: not only thoughts, so communications irrelevant; D: suicide-related behaviors are present; E: outcome is fatal; F: not accidental; G: behavior is not only preparatory; H: without a plan; I: without threats; J: sufficient information): Self-Injurious Behavior, Suicidal
Intent Unknown (C-CASA); Self-Inflicted Death with Undetermined Intent (fatal outcome)  
(Silverman)

A woman has a history of cutting herself on her arms when she is upset. She has talked to her therapist about suicidal thoughts this month, but has never tried to kill herself in the past. A neighbor finds her dead in her apartment, having bled out from a vertical cut to a vein in her arm. There is no suicide note. Her husband of ten years left her for another woman that day.

13. Suicidal ideation (A: without suicidal intent; B: with only thoughts present; C: without communications; D: without suicide-related behaviors; E: without injuries; F: not accidental; G: without preparatory behavior; H: without a plan; I: without threats; J: sufficient information):

Suicidal Ideation (C-CASA); Suicide-Related Ideations, Type I (with no suicidal intent) (Silverman)

After her husband leaves her, a woman thinks about killing herself but decides that she really doesn’t want to die. She never gets to the point of making a plan. She feels embarrassed about thinking about suicide and thus doesn’t tell anyone that she was thinking about killing herself. She is relieved that she doesn’t want to die.

14. Suicidal ideation (A: with unclear suicidal intent; B: only thoughts are present; C: communications are absent; D: communications are absent; E: injuries are absent; F: not accidental; G: without preparatory behaviors; H: without a plan; I: without threats; J: sufficient information): Suicidal Ideation (C-CASA); Suicide-Related Ideations, Type II (with undetermined suicidal intent)

(Silverman)

A woman’s boss criticizes her strongly at work. After work, she goes to a bar and has a lot to drink. While drunk, she thinks about killing herself, though she is so inebriated that her thoughts don’t become specific enough to form a plan. When she ponders her drunken thoughts about suicide later (when sober), she can’t decide whether she really wanted to die or not. She doesn’t tell anyone about her thoughts.
15. *Not Enough Information* (A: suicidal intent is unclear; B: with only thoughts present; C: without communications; D: without suicide-related behaviors; E: without injuries; F: not accidental; G: without preparatory behavior; H: without a plan; I: without threats; J: not enough information):

Not Enough Information (C-CASA); Suicidal Ideation, Type II (with undetermined degree of suicidal intent) (Silverman)

A woman lays in her bed thinking about suicide and self-injury. She does not tell anyone about her thoughts. She does not act on her thoughts. A short time later, she falls asleep.

16. *Suicidal ideation* (A: suicidal intent is present; B: only thoughts are present; C: without communications; D: suicidal behaviors are absent; E: injuries are absent; F: not accidental; G: without preparatory behaviors; H: without a plan; I: without threats; J: sufficient information):

Suicidal Ideation (C-CASA); Suicide-Related Ideations, Type III (with some suicidal intent) (Silverman)

A woman thinks about killing herself. She wishes she were dead but she does not think about particular ways in which she would kill herself. She doesn’t tell anyone about her suicidal thoughts, and she doesn’t act on them. She decides that she intends to kill herself someday, though she lacks the energy to think of a particular plan today.

17. *Suicidal ideation* (A: suicidal intent is absent; B: only thoughts are present; C: with communications present; D: without suicide-related behaviors; E: without injuries; F: not accidental; G: without preparatory behavior; H: without a plan; I: without threats; J: sufficient information): Suicidal Ideation (C-CASA); Suicide-Related Communications, Type I (with no suicidal intent) (Silverman)

A woman is speaking to her therapist about how terrible she feels all the time. She admits that she has suicidal thoughts, but denies having a plan or intent to kill herself anytime in the near future. Her therapist reminds the woman that she always seems to
feel worse in the winter and that spring is coming soon. “You’re right,” the woman said, “I guess I don’t really want to die, I just want to feel better.”

18. **Suicidal ideation (A: with unclear suicidal intent; B: with only thoughts present; C: with communications; D: without suicide-related behavior; E: without injuries; F: not accidental; G: without preparatory behaviors; H: without a plan; I: without threats; J: sufficient information):**

Suicidal Ideation (C-CASA); Suicide-Related Communications, Type II (with undetermined degree of suicidal intent) (Silverman)

A woman who has been diagnosed with Borderline Personality Disorder tells her therapist that she has been thinking about killing herself. “I’m just such a mess,” she states, “I don’t know if I should be around anymore.” The therapist asks her if she has a plan and if she’s thinking about killing himself today. “No, no plan,” the woman replies, “I wouldn’t kill myself today or even this week... in fact, I don’t know if I really mean it or if I just mean that I hate the way I always feel! Maybe I do mean I want to kill myself, maybe I don’t – what do you think, should I kill myself?”

19. **Suicidal ideation (A. suicidal intent is present; B: only thoughts are present; C: communications are present; D: without suicide-related behavior; E: without injuries; F: not accidental; G: without preparatory behavior; H: without a plan; I: without threats; J: sufficient information):** Suicidal Ideation (C-CASA); Suicide-Related Communications, Type III (with some suicidal intent) (Silverman)

A woman has been thinking about killing herself for several months. She never reaches the point where she makes a plan, but the thoughts keep coming into her head. She tells her best friend, “I don’t think it’s going to be today or even this month, but I have a feeling I’m going to end up killing myself someday.” Her friend asks her if she really means that and she replies, “Yes, I really do intend to end my life at some point, I just
think it will be a while from now.” He laughs uncomfortably and tells her to “stop talking crazy!”

20. **Suicidal ideation (A: suicidal intent is absent; B: only thoughts are present; C: communications are present; D: suicide-related behavior is absent; E: injuries are absent; F: not accidental; G: without preparatory behaviors; H: without a plan; I: with threats present; J: sufficient information):**

   Suicidal Ideation (C-CASA); Suicide Threat, Type I (with no suicidal intent) (Silverman)

   A woman calls a suicide hotline. She explains that she is thinking of killing herself. The counselor empathizes with her emotional pain; then he asks if she has a plan. She denies having a plan, stating, “I don’t think I would actually kill myself... I keep thinking about killing myself today, but I might be too weak to actually go through with it. I’m so pathetic.” The counselor tries to clarify whether her recurrent suicidal thoughts would lead her to harm herself today. Eventually she states that she does not intend to kill herself today, even though she reiterates that she keeps thinking about suicide.

21. **Suicidal ideation (A: suicidal intent is unclear; B: only thoughts are present; C: communications are present; D: suicide-related behavior is absent; E: without injuries; F: not accidental; G: without preparatory behavior; H: without a plan; I: with threats present; J: sufficient information):** Suicidal Ideation (C-CASA); Suicide Threat, Type II (with undetermined degree of suicidal intent) (Silverman)

   A woman is out to dinner with a friend. She has had thoughts of killing herself throughout the day but can’t make up her mind whether she wants to die. She suddenly blurts out, “I don’t know if I can go on anymore!” Her friend, shocked, doesn’t know what to say. The woman clarifies, “I feel like I might do something to hurt myself tonight, but I’m not sure. It’s not like I have a plan or anything. I’m scared and I’m tired.” Her friend tells her that he cares about her and he doesn’t want her to kill herself. They awkwardly resume eating.
22. **Suicidal ideation** (A: suicidal intent is present; B: only thoughts are present; C: communications are present; D: without suicide-related behavior; E: injuries are absent; F: not accidental; G: without preparatory behavior; H: without a plan; I: threats are present; J: sufficient information):

Suicidal Ideation (C-CASA); Suicide Threat, Type III (with some suicidal intent) (Silverman)

A woman lays in bed, unable to sleep. She feels sad even though nothing has happened to make her sad. She has felt like this for a long time. Believing that nothing will make her feel better, she thinks about killing herself. She wakes her husband up and says, “I don’t think I can go on any longer. I’m going to kill myself when you’re at work today. I just wanted to say goodbye.” Scared, he asks him how she’s going to kill himself, thinking that maybe he should take her to the hospital. “I don’t know how I’ll do it,” she replies, “Just that I’m not going to be here when you get home.”

23. **Preparatory behavior** (A: suicidal intent is absent; B: not only thoughts; C: not only thoughts, so communications irrelevant; D: suicide-related behaviors are present; E: injuries are absent; F: not accidental; G: behaviors are only preparatory; H: presence of a plan; I: absence of threats; J: sufficient information):

Self-Injurious Behavior, No Suicidal Intent (C-CASA); Suicide Plan, Type I (with no suicidal intent) (Silverman)

A woman goes to the store and purchases a very sharp knife. She intends to cut herself with this special knife when she is feeling down. However, she does not intend to kill herself with it. Her plan is to make small cuts on her arms with the special knife when she next feels hopeless and frustrated. Now, though, she feels okay so she doesn’t cut himself.

24. **Preparatory behavior** (A. suicidal intent is unclear; B: not only thoughts; C: not only thoughts, so communications irrelevant; D: suicide-related behaviors present; E: injuries are absent; F: behavior is not accidental; G: behavior is only preparatory; H: presence of a plan; I: absence of
threats; J: sufficient information): Self-Injurious Behavior, Suicidal Intent Unknown (C-CASA); Suicide Plan, Type II (with undetermined degree of suicidal intent) (Silverman)

A woman goes to the store and purchases a gun so that she could have it available in case she would decide to kill himself. She deliberately did not buy bullets, though, to keep herself safe. She then makes a plan to purchase bullets, but her husband sees her notes and asks her what the bullets would be for. She denies any suicidal intent, but it’s unclear if she’s lying. Because of his interruption, she does not go through with buying bullets.

25. Accident (A: suicidal intent is unclear; B: only thoughts are present; C: with communications; D: without suicide-related behaviors; E: with injuries present; F: it was accidental; G: without preparatory behavior; H: with a plan; I: without threats; J: sufficient information): Other, No Deliberate Self-Harm (C-CASA); Suicide Plan, Type II (with undetermined degree of suicidal intent) (Silverman)

A woman is leaving her therapist’s office after talking about her plan to kill herself next week. Distracted by her thoughts about what her therapist and she were discussing, she misses a step on the stairs and falls down a flight of stairs. She is taken to the hospital for lacerations to her legs and is given x-rays to see if anything is broken.

26. Preparatory behavior (A: suicidal intent is present; B: not only thoughts; C: not only thoughts, so communications irrelevant; D: suicide-related behaviors are present; E: injuries are absent; F: not accidental; G: behavior is only preparatory; H: with a plan; I: without threats; J: sufficient information): Preparatory Acts Toward Imminent Suicidal Behavior (C-CASA); Suicide Plan, Type III (with some suicidal intent) (Silverman)

A woman makes a plan to kill herself today, Saturday, by swallowing pills. She goes around her house collecting pill bottles but leaves them unopened. She puts them all in her medicine cabinet, together in one spot. She intends to swallow them within the
hour, but suddenly realizes that her financial affairs are not yet in order. She closes the

cabinet, disappointed that she has to live long enough to talk to her accountant on

Monday.
APPENDIX C

DEMOGRAPHIC QUESTIONNAIRE

1. How many hours per week are you engaged in psychological service delivery? In this context, psychological service delivery is defined to include: direct client care; providing supervision to others regarding their direct client care; receiving calls at a mental health call center or hotline; receiving supervision regarding direct client care; writing case notes regarding client care; conducting assessments for the purpose of direct client care; writing assessment reports for the purpose of direct client care; preparing case conference materials regarding client care; participating in case conferences regarding client care; and consultation with other providers regarding client care. In short, psychological service delivery includes any activity done for the purpose of promoting direct client care; this excludes research and assessments that are not conducted for the purpose of promoting direct client care (such as academic research about psychological service delivery).
   a. Less than 5 hours per week
   b. 5-10 hours per week
   c. 11-20 hours per week
   d. 21-30 hours per week
   e. 31-40 hours per week
   f. More than 40 hours per week

2. What is your marital status?
   a. Single, never married
   b. Single, in a long-term relationship, or engaged
3. What is your gender?
   a. Male
   b. Female

4. What is your race/ethnicity?
   a. White
   b. Black
   c. Hispanic
   d. Asian
   e. Pacific Islander
   f. American Indian
   g. Alaskan Native
   h. Other (please specify): ________________
   i. Don’t know

5. In what mental health discipline are you licensed?
   a. Psychologist
   b. Marriage and family therapist
   c. Social Worker
   d. Professional counselor
   e. Psychiatric nurse
   f. Psychiatrist
   g. Other (e.g., “Student,” please specify): ________________
6. Please estimate the number of clients you have worked with (during your career) who have had suicide-related thoughts or behaviors: _________________
NOTICE OF APPROVAL

January 24, 2013

Thomas J. Renkin
2908 West Victory Blvd. Apt. A
Burbank, California 91505

From: Sharon McWhorter, IRB Administrator

Re: IRB Number 20130112 "Comparing the Utility and Reliability of Three Current Suicide-Related Nomenclatures"

Thank you for submitting your Exemption Request for the referenced study. Your request was approved on January 23, 2013. The protocol represents minimal risk to subjects and matches the following federal category for exemption:

☐ Exemption 1 – Research conducted in established or commonly accepted educational settings, involving normal educational practices.

☐ Exemption 2 – Research involving the use of educational tests, survey procedures, interview procedures, or observation of public behavior.

☐ Exemption 3 – Research involving the use of educational tests, survey procedures, interview procedures, or observation of public behavior not exempt under category 2, but subjects are elected or appointed public officials or candidates for public office.

☐ Exemption 4 – Research involving the collection or study of existing data, documents, records, pathological specimens, or diagnostic specimens.

☐ Exemption 5 – Research and demonstration projects conducted by or subject to the approval of department or agency heads, and which are designed to study, evaluate, or otherwise examine public programs or benefits.

☐ Exemption 6 – Taste and food quality evaluation and consumer acceptance studies.

Annual continuation applications are not required for exempt projects. If you make changes to the study’s design or procedures that increase the risk to subjects or include activities that do not fall within the approved exemption category, please contact me to discuss whether or not a new application must be submitted. Any such changes or modifications must be reviewed and approved by the IRB prior to implementation.

Please retain this letter for your files. This office will hold your exemption application for a period of three years from the approval date. If you wish to continue this protocol beyond this period, you will need to submit another Exemption Request. If the research is being conducted for a master’s thesis or doctoral dissertation, the student must file a copy of this letter with the thesis or dissertation.

Cc: Ronald F. Levant - Advisor
    Valerie Callanan - IRB Chair

☐ Approved consent form/s enclosed

The University of Akron is an Equal Educator and Employment Institution

153
Informed Consent

Purpose: The purpose of this research is to advance the study of suicide by collecting data that will permit comparison of three major current suicide-related nomenclatures.

Procedures: This research concerns suicide-related nomenclatures: that is, terms and definitions having to do with suicide, suicide attempts, suicidal thoughts, and the like. You will be asked to read about one set of terms and definitions having to do with suicide and then to apply those terms and definitions to 10 brief hypothetical vignettes. The 10 vignettes present a short scenario (about 3-6 sentences) and you will be asked which term and definition would best apply to that scenario. Following the vignettes, you will be asked to answer some questions about your experience of applying the terms and definitions to the vignettes, as well as whether you think the nomenclature would be helpful to clinical practice. At the outset of the research, you will also be asked some demographic questions. It is estimated that your participation in this research should take between 30 and 45 minutes.

Exclusion: This research is intended for mental health clinicians who provide at least 5 hours of psychological service delivery each week. The first survey question will define what psychological service delivery means in this context. If you do not deliver at least 5 hours of psychological services each week, you will be thanked for your participation and the software will end the survey for you.

Risks and Discomforts: This research is expected to have minimal risk of harm to participants. One risk might be if thinking about suicidal thoughts or behaviors triggered your own mental health issues; or if you have experienced the death of a loved one by suicide, then reading the vignettes might remind you of your grief. In the unlikely event that you should have suicidal thoughts or contemplate self-harming behaviors during the study, please call the national suicide prevention lifeline at 1-800-273-TALK (8255). Their website is available at www.suicidepreventionlifeline.org.

Benefits: The benefits of participation include the opportunity to think more deeply about the language we use to describe suicide-related thoughts and behaviors. In addition, participants will be entered into a random drawing for one $200 gift card to Amazon.com. Since this research seeks the participation of 180 participants, your chances of winning the gift card are about 1 in 150. Participants will need to provide their e-mail address to be eligible to win the gift card.

Right to Refuse or Withdraw: As with any research, you have the right to discontinue your participation at any time and for any reason. However, this research depends on participants' complete answers to all survey questions, so please make a good faith effort to complete the survey in its entirety unless the experience is actually causing you emotional or other distress.

Confidentiality of Records: All data will be kept in password-protected computers for at least three years, as required by federal law. At that point, aggregate data may be retained securely for analysis in future research, but individual data may be destroyed. No individual-level data will be reported in Mr. Rankin's dissertation itself. Your responses will be kept entirely confidential.

*Informed Consent Signature: Since participants will not be asked to disclose any personal information about themselves (except demographics) during the survey, we ask that you sign this informed consent form by typing your first and last name below. By doing so, you acknowledge that you understand the risks and benefits of participating in this research and that you freely choose to participate.

Signature (type your first and last name):