The Validation of A Measure of Competency in the Use of Psychological Assessment in Career Counseling: A Piagetian Framework

Roy L. Etheridge
FLORIDA STATE UNIVERSITY
COLLEGE OF EDUCATION

THE VALIDATION OF A MEASURE OF COMPETENCY IN THE USE OF
PSYCHOLOGICAL ASSESSMENT IN CAREER COUNSELING:
A PIAGETIAN FRAMEWORK

By
ROY L. ETHERIDGE

A Dissertation submitted to the
Department of Educational Psychology and Learning Systems
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Degree Awarded:
Summer Semester, 2008
The members of the Committee approved the dissertation of Roy L. Etheridge, defended on July 26, 2007.

______________________________
Gary W. Peterson
Major Professor, Dissertation Director

______________________________
Richard L. Tate
Outside Committee Member

______________________________
Briley Proctor
Committee Member

______________________________
R William English
Committee Member

Approved:

______________________________
Aki Kamata, Ph.D., Department Chair,
Department of Educational Psychology and Learning Systems

The Office of Graduate Studies has verified and approved the above named committee members.
This dissertation is dedicated to my brother, David A. Etheridge. He regularly reminded me that
the people we become are forged by our actions more than our words. A man of few words, he
did many kind things for others without a word being uttered. Moreover, he taught me that things
in life are not supposed to be anything other than what they are. I hope to carry his values in my
actions and words for as long as I live.
ACKNOWLEDGEMENTS

I would like to thank my wife, Dr. Mary Anne Etheridge, for her invaluable assistance as editor of this dissertation and chief of our household. A special thanks is extended to my major professor, Dr. Gary Peterson, for his patience and guidance during the last 8 years. Additional thanks are extended to each member of my doctoral committee. All of them have participated in my personal and professional development. Thank you for your patience, understanding, and continued belief in me.

I wish to acknowledge the efforts of the expert and student participants of this study. The performance tasks for all parties were significant and are worth mention in this section. Finally, I wish to acknowledge Dennis Carroccio and Darran Duchene. Both have acted in a supervisory capacity at my place of employment over the last six years. Their support of my family and my goals will always be appreciated.
# TABLE OF CONTENTS

List of Tables ............................................................................................................... viii
Abstract ......................................................................................................................... ix

1: INTRODUCTION ..................................................................................................... 1
   Social Problem .................................................................................................... 1
   Professional Problem........................................................................................... 1
   Theory Base ........................................................................................................ 2
   Purpose................................................................................................................ 5
   Definitions of Terms............................................................................................ 5
   Delimitations ....................................................................................................... 6
   Assumptions ........................................................................................................ 7
   Significance of the Study ..................................................................................... 7

2: REVIEW OF THE LITERATURE.............................................................................. 8
   The Need for a Measure of Competency.............................................................. 8
   Recent Efforts in Assessing Counselor Trainee Competence................................ 9
   The History of Assessor Competency Measurement .......................................... 10
   Qualifications .................................................................................................... 12
   The Test Users Qualifications Work Group (TUQWoG).................................... 12
   Task Force on Test User Qualifications (TFTUQ) ............................................. 15
   Generic Qualifications ....................................................................................... 15
   Specific Qualifications....................................................................................... 16
   Specific Qualifications: Career/Vocational Counseling Context......................... 17
      Career Context: Classification ................................................................ 18
      Career Context: Description ................................................................... 18
      Career Context: Prediction ..................................................................... 18
      Career Context: Intervention Planning .................................................... 19
      Career Context: Tracking ....................................................................... 20
   Qualifications: Summary ................................................................................... 20
   Rationale for Developing a New Model of Competency Assessment .................. 22
   Competency and Counselor Development.......................................................... 22
   Piaget .............................................................................................................. 24
   Developing a Measure of Competency. ............................................................... 25
   Summary of the Matrix Formation and the Next Step ........................................ 26
   Building the Instrument ..................................................................................... 26
   Process Mapping: Steps toward Exemplary Test Use ......................................... 27
   The Next Step: Validating the Matrix of Assessor Competency......................... 29

3: METHODOLOGY .................................................................................................... 30
   Study 1: Content Validity .................................................................................. 30
      Purpose ......................................................................................................... 30
LIST OF TABLES

Table 1. Participant Characteristics ................................................................. 34
Table 2. Pearson Correlations ........................................................................... 43
Table 3. Expert Ratings of the Importance of AMAC-CTI Items .................... 45
Table 4. Interrater Intra-Class Correlations for the Overall Score ................. 45
Table 5. Interrater Intra-Class Correlations for Item 1: Purpose ..................... 46
Table 6. Interrater Intra-class Correlations for Item 2: Constructs ................. 47
Table 7. Interrater Intra-class Correlations for Item 3: Use ............................ 47
Table 8. Interrater Intra-class Correlations for Item 4: Interpretation .......... 48
Table 9. Interrater Intra-class Correlations for Item 5: ILP ............................. 48
Table 10. Interrater Intra-class Correlations for Item 6: Intervention .......... 49
Table 11. Assessment of AMAC-CTI Items ..................................................... 50
Table 12. Eigen Values and Percent Variance ............................................... 50
Table 13. Descriptive Statistics Regarding Task Performance ....................... 51
Table 14. Total Ratings of Zero ....................................................................... 52
Table 15. Total Ratings of Zero (n=18) .......................................................... 53
Table 16. Comparison of Student Overall AMAC-CTI Ratings and Supervisor Ratings 54
Table 17. Correlation: Education, Experience, and AMAC-CTI Scores (n = 33)........ 55
Table 18. Correlation and Descriptive Statistics: Square Root Experience (n = 33)..... 56
Table A1. Piagetian Matrix of Test User Competence (Theoretical Model) ....... 68
Table A2. Competency Analysis ..................................................................... 93
Table A3. Individual Results .......................................................................... 93
ABSTRACT

Based on the results of a prior field study, it was determined that an instrument that borrows from Jean Piaget’s theory of cognitive development could be constructed and reliably used to measure assessor competence in the use of a career counseling assessment instrument in career counselor training. This research further explores the feasibility of validating this instrument. If successful, a training template could be created to provide competency measurement for the remediation of counselors in training and the improvement of counselor training models. The theoretical model upon which the instrument is based is the Piagetian Matrix of Test User Competence (PMTUC). The competency assessment instrument based on this theoretical matrix was named A Measure of Assessor Competence (AMAC). The AMAC produces one global score based on six test items. The long-term intent of this line of research is to promote the utility of the PMTUC in the creation of a variety of measures of competency (AMACs) across many psychological assessments. The PMTUC theory and the resulting AMAC instruments could be applicable to all instruments. The specific intent of this study was to validate the use of the AMAC in the creation of a measure of competency in the use of a career counseling instrument. The instrument selected for this validation research was the Career Thoughts Inventory (CTI) because experts in the use of this instrument were readily available. Therefore, the measure of assessor competency for this specific research study is the AMAC-CTI. Future studies might attempt to build measures of competency in the use of the MMPI-2 (AMAC-MMPI-2), Rorschach (AMAC-Rorschach), or perhaps the WISC-IV (AMAC-WISC-IV).

To validate the AMAC-CTI instrument, five studies were conducted. Study 1 involved expert ratings of the importance of the six items in the AMAC-CTI using an Expert Content Rating Form. The experts in the field of counseling and career development were identified by the Director of Clinical Training in a Combined Doctoral Program in Counseling Psychology and School Psychology at a large university in the southeastern United States. For this study, persons were considered experts if they had at least 10 years experience in the field of career counseling, held faculty positions, currently supervised graduate students in career counseling, and had served as a supervisor for the student administration of at least two hundred Career Thoughts Inventories. All five experts agreed that the items on the AMAC-CTI were important
Studies 2, 3, and 4 involved expert raters, graduate students, and professionals in the field of counseling and career development. The graduate students were enrolled in a Combined Counseling Psychology and School Psychology doctoral program or the Mental Health Counseling masters program at a large southeastern university who have been trained in the use of the CTI. The professionals work in the field of counseling psychology and have also been trained in the use of the CTI. Participants were approached via face-to-face request, e-mail request, or telephone by either the primary investigator of this dissertation or the aforementioned Director of Clinical Training about volunteering for a study of trainee competency using assessments. Once persons agreed to participate, they were contacted via e-mail by the primary investigator and were directed via e-mail to access a web link provided by www.surveymonkey.com. Once participants accessed the link, they were introduced to the survey and presented with an electronic consent form and, upon agreeing to participate, a background questionnaire. Participants provided responded to six open-ended format questions which were assumed to correspond to the 6 primary determinants of test user competence. At the conclusion of the survey collection process, responses to surveys were redacted of personal identification information and given to expert raters to perform ratings using the AMAC-CTI.

For Study 2, inter-rater reliability coefficients and measures of internal consistency were derived to confirm the reliability of the instrument. An exploratory factor analysis (EFA) determined that the AMAC-CTI is a uni-dimensional instrument. Study 3 was conducted to examine the difficulty of the instrument. The open-ended portion of the survey required respondents to answer six detailed questions that corresponded to the six items that make up the AMAC-CTI. Based on the results of this research project, the performance tasks were determined to be somewhat difficult.

Study 4 assessed convergent validity by asking the student participants’ clinical supervisors to rate their respective students’ competency in the use of the CTI. Supervisors used the same evaluation criteria as the AMAC-CTI to assess their students. The student participants’ overall AMAC-CTI scores were then correlated with the overall ratings provided by their respective clinical supervisors. It was hypothesized that these scores would be correlated, but statistical analyses failed to show a significant relationship. For Study 5, analyses were performed to examine the relationship between AMAC-CTI scores and education and between
AMAC-CTI scores and experience in the use of the CTI. AMAC-CTI ratings were positively correlated with experience in the use of the CTI, but were not correlated with education level and the number of assessment courses completed by participants. Implications for further test development and counselor training of assessment skills are discussed.
CHAPTER 1
INTRODUCTION

This chapter begins with the social and professional problems that are the rationale for investigation into this topic. It presents an introduction into the theory base as well as the purpose of this research. Definitions of relevant terms are included for reference. The delimitations and assumptions of this study are presented, as well as the study’s significance.

Social Problem

Counselors must be competent in the proper selection, administration, interpretation, and application into practice of the assessment instruments they use in counseling. The majority of the governing bodies of practicing psychologists have promulgated rules and standards pertaining to the appropriate use of psychological assessments (American Association for Marriage and Family Therapy, 1985; American Counseling Association, 1995; American Educational Research Association, 1985; American Psychological Association, 1999; National Association of School Psychologists, 1992; National Council on Measurement in Education, 1995). The original emphasis of test practices focused on psychometric knowledge. Recent standards emphasize the need for test users to be skilled and knowledgeable about specific contexts and rationales of use. While most training programs insist that students familiarize themselves with the codes and standards of practice, a measure to determine the rationale or intended use of a given instrument is currently not readily available. Regardless of the intent of the governing bodies above, an outline of standards is only as effective as the tools available to measure the actions of individuals supposedly governed by those standards (Etheridge, Peterson, & Hill, 2004). The goal of this project was to validate a measure that gauges the level of competency counselors demonstrate when using a given assessment tool.

Professional Problem

The American Psychological Association (APA), American Educational Research Association (AERA), and National Council on Measurement in Education (NCME) joined to form the Joint Committee on Test Use Practices (JCTP). The JCTP was created to address
growing concerns about test user competence. This group recently published a revision of their Standards for Educational and Psychological Testing (1999). A review of these revised standards suggests that these three governing bodies are committed to outlining the most comprehensive standards possible regarding test use. Borrowing from standard 11.4 (1999), test users should “have a clear rationale for the intended uses of a test or evaluation procedure in terms of its validity and contribution to the assessment and decision making process” (p. 114). This study was specifically designed to address the degree of trainee competency in decisions to select a given test. Previous research and competency standards assessed the accuracy trainees exhibited in test selection, but few appear to have assessed the specific rationale that trainees use when deciding to use a specific test for a specific client.

A great deal of research has been committed to comparing novice and expert counseling techniques (Glaser & Chi, 1988; Hiebert, Hallberg, & Cummings, 1989; Hillerbrand, 1987; Hillerbrand & Claiborne, 1990; Mayfield, Kardash, & Kivlighan, 1999; Martin, Slemon, Kivlighan & Quigley, 1991), developmental changes in trainees (Heppner & Roehlke, 1984; Reising & Daniels, 1983; Stoltenberg, 1981; Thompson, 1986; Tracey, Hays, Malone, & Herman, 1988; Zimpfer, 1996; Zimpfer & Detrude, 1990), and the personal characteristics of effective counselors (Jackson & Thompson, 1971; Reising & Daniels, 1983; Thompson, 1986; Wiggins & Weslander, 1979). These studies provide the profession with a greater understanding of the student counselor, the relative effectiveness of student counselors, and how novice counselors become experts, but they fail to address the level of competency exhibited by the student counselor.

The majority of these studies have found it difficult to operationalize the meaning of the term “expert” or “expertise” (Hillerbrand & Claiborne, 1990; Jennings & Skovholt, 1999; Martin et al., 1989; Orlinsky, 1999). Given this difficulty, another possible solution was to focus on the most recent standards of test use promulgated by the APA and develop a competency model based on standards outlined by the governing body of our profession, psychology. This was the objective of the following endeavor.

**Theory Base**

The American Psychological Association (APA) was one of the first organizations to be concerned with test user qualifications. In the late 1940s, the APA formed the Committee on
Ethical Standards for Psychology and one of the first topics this committee tackled was the sale and distribution of psychological tests and diagnostic aids (APA, 2000; Hobbs, 1951). In 1950, the committee released its first set of ethical standards regarding the distribution of tests. The complete set of standards was released three years later in 1953 (Golann, 1970). Since 1953, a number of organizations have developed standards and promulgated ethical guidelines regarding the use of psychological tests (American Association for Marriage and Family, 1998; American Counseling Association, 1988; American Educational Research Association, 1999; National Association of School Psychologists, 1992; National Council on Measurement in Education, 1995). In 2000, the APA’s Task Force on Test User Qualifications issued the most comprehensive standards of test use to date (Turner, DeMers, Fox, & Reed, 2001).

In tracking the progression of events to date, the APA set out in 1950 to establish guidelines that would assist test publishers in the distribution of psychological assessments. Thirty years later, the APA acknowledged that the system they had developed in 1950 was outdated. Testing had become much larger than any of the APA’s forefathers could have imagined. In 1984, the APA merged with the AERA and NCME and developed a Joint Committee of Testing Practices. This was a merger of several associations committed to establishing more coherent guidelines for test use. From the JCTP, a work group was created to develop a data-based approach to promoting good test use. Again, the emphasis of this workgroup was to develop guidelines that would assist test publishers in the dissemination of testing materials.

Shortly after the work group completed their task, the APA acknowledged that test misuse was not entirely tied to publishers and their respective distribution practices. In fact, they admitted that it was likely that the majority of test misuse was tied to the misguided or misinformed practices of test users (APA, 2000; Turner et al., 2001). In 1996, the APA assembled the Task Force on Test User Qualifications (TFTUQ). The TFTUQ was established because the APA believed there was evidence that “some current users of psychological tests may not possess the knowledge and skill that the APA considers desirable for maximum test use” (APA, 2000, p. 6). The task force borrowed from the work of the TUQWoG and developed two global sets of test user qualifications: generic and specific.

Generic qualifications are basic competencies, knowledge, or skills that all test users should possess. These competencies include basic knowledge about statistics, measurement, and
diverse populations. Specific qualifications are best described as five basic competencies that are applicable across different practice contexts. These qualifications vary according to context, and test users should possess specific knowledge in specific areas when using assessments in a given context. The five competencies include classification, description, prediction, intervention planning, and tracking. These specific competencies include specific guidelines about when a test is used, why a test is selected, to whom a test is administered, how results are interpreted, and what is shared with clients. The five contexts that these competencies apply to in varying degrees include the employment context, educational context, career/vocational counseling context, health care context, and forensic context.

In the current study, the career/vocational context and the competencies relative to this domain were explored. The career/vocational counseling contextual guidelines emphasize the importance of properly selecting instruments, differentiating which assessments are appropriate for which clients given their developmental level, and differentiating between the most important and developmentally appropriate forms of information that should be shared with clients. Given this, it is essential that career/vocational counselors know why they select the tests they do, when to administer tests, what information they share with clients, and how they integrate information gained from tests in a meaningful and competent manner.

One might imagine that a historical overview of the standards of practice, ethical conduct of clinical assessment, and the history of test use practices would be much larger in scope than was presented above. The truth is that research in this area has been minimal until recent years. While this may seem illogical, it makes sense after one attempts to operationalize concepts like abstract thinking and the assessment of abstract processing in trainee use of tests. Content domains are much easier to assess and make up the majority of the competency testing in the field of psychology. Several of the authors referenced thus far have expressed a great deal of concern about counselor trainee competency. They argue that counselor training programs fail to adequately prepare future clinicians for clinical assessment (APA, 2000; Baldo et al., 1997; Custer, 1994; Eyde et al., 1988; Frame & Stevens-Smith, 1995; Knoff & Prout, 1985; Lumadee & Duffey, 1999; Meyer, 1980; Robertson & Eyde, 1986). Moreover, nearly all of the authors cited above reported being surprised by the paucity of literature in this area. Based on the summary above, guidelines for test distribution and competency guidelines for test use have been comprehensively delineated. To date, however, there are no tools available to determine if the
guidelines put forth are being met. Clearly, the next step is the development of a competency assessment model that can measure the abstract level in which trainees are operating when using tests: a model designed to assess the specific competencies outlined by the American Psychological Association (2000). As stated earlier in this review, any set of guidelines, qualifications, or competencies is only as effective as the measures used to verify their proper use.

**Purpose**

The primary purpose of this dissertation was to develop a reliable and valid measure of assessor competency. The theory upon which this measure was based is Piaget’s theory of cognitive development. It is also based on theoretical models that suggest competency and professional growth can be accurately measured. The AMAC-CTI was developed as a prototype of competency measurement that attempts to quantify thinking processes in a meaningful and consistent manner. This study sought to demonstrate the reliability and validity of the AMAC-CTI instrument. The validation techniques included in this study were criterion, content, and concurrent validity, internal consistency, and inter-rater reliability.

**Definitions of Terms**

1. **Convergent validity**: The degree to which a measure predicts some criterion. For this study, it examined whether the AMAC-CTI scores are related to education and experience.
2. **Content Validity**: The degree to which test content is tied to the instructional domain it intends to measure. For this study, experts examined and rated the relevance of the 6 domains of the AMAC-CTI.
3. **Convergent Validity**: The degree to which a measure yields a result consistent with the results of another measure of the same construct. For this study, supervisory ratings of participants across the 6 domains of the AMAC-CTI were compared to AMAC-CTI ratings.
4. **Concurrent Validity**: The degree to which a measure yields a result consistent with the results of another measure of the same construct. Concurrent validity is different from convergent validity because the other measure used for comparison must be well-validated.
5. **Inter-rater Reliability**: The degree to which the measuring instrument yields similar results at the same time with more than one assessor. For this study, three expert raters rated subject responses using the AMAC-CTI. These ratings were compared to determine degree of agreement.

6. **Internal Consistency**: Estimation based on the correlation among the variables comprising the set.

7. **Assessor Competence**: The proper selection, administration, and interpretation of results for a given psychological assessment.

8. **Concrete Operational Stage**: A stage of development during which individuals organize ideas in a rudimentary manner along a single dimension, such as size. For this study, the concrete stage was characterized by counselor decisions based on contextual criteria (e.g., when a client is in career distress, use a career counseling instrument).

9. **Transitional Operational Stage**: A stage of development during which individuals are capable of abstract propositions and some testing of hypotheses. For this study, transitional operational referred to counselor decisions based on contextualized and decontextualized cues (e.g., when a client is in career distress, use a career instrument that is unique to the client’s specific needs).

10. **Formal Operational Stage**: A stage of development during which thought processes are highly logical allowing for abstract propositions and testing of hypotheses. For this study, counselor decisions based on decontextualized cues and knowledge base were considered to be at the formal operational stage (e.g., when a client is in career distress, use career instrument that is specific to client and that will link to treatment plan in a meaningful manner for each client).

11. **Psychological Assessment**: A measure of client need, distress, or character.

12. **Competence**: A real and demonstrated ability to successfully carry out a given task. For this study, it was the degree to which participants exhibited appropriate decision making regarding the selection, administration, and interpretation of the CTI.

13. **Qualification**: A skill or set of skills necessary to achieve competence.

**Delimitations**

There were several limitations to this study. The sample size was small. The only
psychological assessment used was the Career Thoughts Inventory. The AMAC-CTI consists of only six test domains and three levels. Subject participation was voluntary and all of the participants were from one graduate program in the southeastern United States. The accurate validation of this instrument was highly contingent upon respondents providing effortful responses to the six survey questions designed to measure their level of understanding.

Assumptions

This investigator identified nine primary assumptions, the most notable of which was the assumption that participants would provide an effortful response to the six questions in the AMAC-CTI. It was also assumed that levels of mastery can be measured, the AMAC-CTI adequately covers the skills linked to competency, and assessor skills are stage-like where each stage builds on earlier stages and the sequence is invariant. It was further assumed that each component of one’s global competency develops independently, some components may be more developed than others, the answers that participants provide reflect their level of development, and the questions are phrased in a manner that evokes responses that reflect the true depth of participants’ understanding.

Significance of the Study

The purpose of this endeavor was to develop a theoretically based model to assess the degree of competency displayed in the use of the Career Thoughts Inventory in career counseling. In so doing, this model could be used for training and evaluation purposes regarding the use of CTI and with a multitude of other instruments beyond the CTI. More importantly, this model could be used to insure that the competencies put forth by the APA and ACA are being met by current test users and test users to come. This research addresses a need to translate professional standards into theory and practice.
This chapter begins with an argument regarding the need for a measure of assessor competency. This is followed by a detailed history of assessor competency measurement and the qualifications that the professional bodies of the counseling profession have promulgated in an effort to promote assessor competence. The chapter ends by examining the role that cognition plays in counselor development and describing a process by which this development can be measured.

The Need for a Measure of Competency

APA, AERA, and NCME joined to form the Joint Committee on Test Use Practices (JCTP). The JCTP was created to address growing concerns about test user competence. This group recently published a revision of their Standards for Educational and Psychological Testing (1999). A review of these revised standards suggests that these three governing bodies are committed to outlining the most comprehensive standards possible regarding test use. Similar to the ACA codes and standards of practice, these standards outline content and process competencies that should be followed verbatim. Borrowing from standard 11.4 (1999), test users should have a clear rationale for the use of tests and this rationale should be based on the validity and understood contribution of the assessment. This study was specifically designed to address the degree of trainee competency in decisions to select a given test. Previous research and competency standards assessed the accuracy trainees exhibited in test selection, but few appear to have assessed the specific rationale trainee’s use when deciding to use a specific test for a specific client.

Unfortunately, a measure to determine the rationale or intended use of a given test is not readily available. Regardless of the intent of the ACA or APA, an outline of standards is only as effective as the tools available to measure the actions of individuals supposedly governed by those standards. More directly, no matter how comprehensive a law of appropriate behavior might be, if there is no measure of adherence, the profession is forced to guess and/or hope that it is being employed. The primary intent of this research was to develop an instrument that might
measure the abstract processing (of the clinician/test user) that is inherent to a standard like the aforementioned Standard 11.4.

Recent Efforts in Assessing Counselor Trainee Competence

A plethora of research has been committed to comparing novice and expert counseling techniques (Glaser & Chi, 1988; Hillerbrand, 1987; Hillerbrand & Claiborne, 1990; Kivilighan and Quigley, 1991; Martin et al., 1989; Mayfield et al., 1999), developmental changes in trainees (Heppner & Roehlke, 1984; Leach & Stoltenberg, 1997; Pope et al., 1976; Reising & Daniels, 1983; Stoltenberg, 1981; Stoltenberg, 2005; Stoltenberg, McNeil, & Crethar, 1994; Thompson, 1986; Tracey et al., 1988; Zimpfer, 1996; Zimpfer & Detrude, 1990), and the personal characteristics of effective counselors (Jackson & Thompson, 1971; Reising & Daniels, 1983; Thompson, 1986; Wiggins & Weslander, 1979). These research studies provide the profession with a greater understanding of the student counselor, the relative effectiveness of student counselors, and how novice counselors become experts, but fail to address the level of competency exhibited by the student counselor.

While counselor training programs have sought to emulate the standards mandated by professional education, there are no formulated theoretical models and operational definitions related to the level of cognitive performance exhibited by trainees that insure these standards are being applied. More directly, there are no measures currently being used that insure that trainees are selecting tests or assessments based on a “clear rationale”. This is not to imply that students are erroneously administering assessments and knowingly circumventing the standards outlined by the governing agencies of their profession. Rather, this is an argument that suggests that the profession of psychology does not know the standards. The only widely used measures of student knowledge are the content-based items included in national and state professional certification and licensing examinations. The respective item pools often contain content questions related to reliability and validity but rarely do they include the domains of selection, administration, interpretation, and application.

Several gatekeeper models have recently been proposed to assess the subsidiary standards of trainee progress (Baldo, 1997; Custer, 1994; Frame & Stevens-Smith, 1995; Lumadee & Duffey, 1999), but none of these models address test use or user competence. Theorists have explored the competency of test use practices (Elmore et al., 1993; Tymofievich & Leroux,
but with little to no emphasis on building models of competency evaluation beyond regular in-service training. As noted by Lumadee and Duffey (1999), there are far too few instruments that measure competency given the nature of the counseling profession.

The Lumadee and Duffey (1999) gatekeeper model directs us toward the idea of models that explore the importance of expanding counselor training beyond academic performance. They contend one of the best ways to do this seems to be by comparing the skills of “expert” versus “novice” counselors. This comparison could help in the development of models that cultivate “expert” techniques. Moreover, the gatekeeper model would likely promote “expert-like” behavior in trainees and thus, show marked improvement in the ways in which neophytes process information in applied training paradigms. However, the majority of these studies have found it difficult to operationalize the meaning of the terms “expert” or “expertise” (Hillerbrand & Claiborne, 1990; Martin et al., 1989; Jennings & Skovholt, 1999; Orlinsky, 1999). Orlinsky’s (1999) commentary on the difficulty of defining the term expert captures the problem most eloquently when he asks, “The master therapist: ideal character or clinical fiction?” Given this difficulty, another possible solution is to focus on the most recent standards of test use promulgated by the APA and develop a competency model based on standards outlined by the governing body of our profession, psychology. This was the objective of the following endeavor.

The History of Assessor Competency Measurement

The American Psychological Association (APA) was one of the first organizations to be concerned with test user qualifications. In the late 1940s, the APA formed the Committee on Ethical Standards for Psychology and one of the first topics this committee tackled was the sale and distribution of psychological tests and diagnostic aids (APA, 2000; Hobbs, 1951). In 1950, the committee released its first set of ethical standards regarding the distribution of tests. The complete set of standards was released three years later in 1953 (Golann, 1970). Since 1953, a number of organizations have developed standards and promulgated ethical guidelines regarding the use of psychological tests (American Association for Marriage and Family, 1998; American Counseling Association, 1988; American Educational Research Association, 1999; National Association of School Psychologists, 1992; National Council on Measurement in Education, 1995). In 2000, the APA’s Task Force on Test User Qualifications issued the most comprehensive standards of test use to date (Turner et al., 2001).
The 1950 APA standards used a tier rating system to determine test user qualifications. Level A and B test users were essentially paraprofessionals. Level C users were required to possess at a minimum a Master’s degree, discernable knowledge in testing, and at least one year of supervised experience. This system was abandoned in 1974 because it was difficult to regulate test use practices once testing material was distributed. From 1974 to 1993, test use abuses were largely attributed to the unregulated distribution practices of publishers (Eyde et al., 1988; Robertson & Eyde, 1986; Tyler, 1986). Since 1993, concern has shifted from unethical publishers to incompetent clinicians (APA, 2000; Elmore et al., 1993; Tymofievich & Leroux, 2000). Most notably, concern has grown regarding the gaps in the training of future test users in graduate training programs and the competency problems tied to training gaps in general (Baldo et al., 1997; Custer, 1994; Frame & Stevens-Smith, 1995; Lumadee & Duffey, 1999; Meyer, 1980; Zimpfer, 1996).

Historically, test user qualifications have been “broadly” defined (APA, 2000; Baldo et al., 1997; Eyde et al., 1988; Knoff & Prout, 1985). This is evident in the APA’s revisions of its standards in 1974 and 1985. The 1985 revision still defined user qualifications loosely, and clinical competency was arguably ignored. However, there are clear differences between the 1985 revision and 1999 revision. The 1999 revision is much more specific about various guidelines and testing issues tied to appropriate test use and competency. The changes that are evident in the 1999 revision are products of the Joint Committee on Testing Practices (JCTP).

The JCTP was founded in 1985 and originally included members from the AERA, APA, NCME, and test publishers. Since 1985, the sponsorship listing of the JCTP has continued to grow. The JCTP is tied to a number of diverse associations and organizations that all recognize the need for the development and promulgation of test user qualifications. The JCTP traces its roots to a 1984 meeting involving the three primary agencies above (APA, AERA, and NCME) and test publishers.

As noted in the opening, since 1950, problems with test use abuses have been largely attributed to publishers (APA, 2000; Aiken et al., 1990; Hobbs, 1951). The JCTP established the Test User Qualification Work Group (TUQWoG) primarily to develop guidelines that would assist test publishers in the distribution of test materials and secondly to assist graduate programs in the development of trainee test user skills (Eyde et al., 1988). Again, the TUQWoG was really established in an effort to combat the suspect practices of test publishers. Publishers routinely
disseminated testing materials using outdated criteria and unreliable verification procedures (e.g. forward tests to entities that simply used official letterhead). The workgroup published its executive summary in December 1988 (Eyde et al., 1988). This summary included definitions that greatly assisted test publishers and worked to assist in the development of test user competency as well.

Eight years later, in 1996, the APA created the Task Force on Test User Qualifications (TFTUQ). The TFTUQ was established because evidence suggested some current users of psychological tests might not possess the knowledge and skill that the APA considers desirable for maximum test use (APA, 2000). In essence, the TFTUQ was primarily tasked to deal with the problems associated with graduate program curriculum, clinical competency, and related test user issues. Recently, the focus has shifted from test publishers to test users (AERA, APA, NCME, 1999; APA, 2000). The TFTUQ recently published its “Guidelines for Test User Qualifications (APA, 2000).”

**Qualifications**

The following sections describe the progressive construction of qualifications in the field of counseling. The first two sections contain descriptions of the Test User Qualification Work Group and the Task Force on Test User Qualifications. The Test User Qualification Work Group developed a model test user qualification system (Eyde et al., 1988; Moreland et al., 1995). The Task Force on Test User Qualifications (TFTUQ) developed guidelines that informed test users and the general public of the qualifications that the APA considers important for the competent and responsible use of psychological tests. The remaining eight sections describe the career/vocational counseling context and the specific qualifications outlined by the APA’s guidelines.

**Test User Qualifications Work Group (TUQWoG)**

The TUQWoG’s mission was to develop a model test user qualification system (Eyde et al., 1988; Moreland et al., 1995). This system was to be based on scientific methodology and to clearly outline the competencies necessary for test use. Moreover, the system they developed was to be applicable to all occupations. Thus, a general data based test purchaser form could be developed and employed by publishers so they could screen would be customers using a
standardized form recognized by the APA and related governing bodies. In developing their model, the work group carried out a large-scale research project designed to identify “good testing practices”. Therefore, the initial emphasis of the project was to identify practices, conditions, and behaviors that resulted in test misuse. Despite the emphasis on individual test user competencies, the main impetus of this project was to design a model that outlined competencies that publishers could use to discern what constitutes purchaser entitlement.

The first study involved identifying the content domain of the behaviors linked to poor testing practices or test misuse. This was accomplished by assuming that the process of using tests could be considered a job for which qualifications could be established. Two systematic job analysis methods were employed: Flanagan’s (1954) critical incident method and Primoff’s job element method (Primoff, 1975; Primoff & Eyde, 1988). Both methods use experts in a given domain to qualify and quantify behaviors. Experts were selected by the workgroup based on their expertise in the development and use of tests.

Initially, critical incidents were gathered from test experts via mail surveys. The experts were asked to describe examples of test misuse that they had observed and the circumstances that were linked to the problems they observed. The descriptions that the experts forwarded were transformed into critical incidents. This transformation was accomplished by rephrasing poor testing behaviors into positive test use practices. These “positive” test use practices were further divided into four sub-elements: knowledge, skills, abilities, and other characteristics related to good test use. Eighty-six critical incidents were selected and renamed “Generic Test User Behaviors”. After 86 critical incidents were captured, the job element method was used to quantify the incidents.

The job element method (Primoff & Eyde, 1988) involves collecting observed behaviors that contribute to job success. The behaviors are rated based on 4 questions:

1. Does the behavior characterize superior performance?
2. Do even barely acceptable performers have this behavior?
3. Is it practical to examine this behavior?
4. Is trouble likely if this behavior is not considered in examining?

Nineteen test use experts rated the 86 generic test user behaviors identified in the first study using the job element questions above. The final product of these two studies represents the behaviors that experts deem important in test use and a rating of the behaviors in terms of
their relative importance.

Study 3 was conducted to evaluate and describe the common factors of test misuse for empirically developed clusters of tests based on the ratings above (Eyde et al., 1988). The clusters of test include: Group Educational, Ability and Preference, Learning Disability and Neuropsychological, Individual Intelligence, Readiness, Objective Personality, and Projective Personality tests. The Self-Directed Search failed to fall into any of the aforementioned categories and was identified as a cluster on its own. Study 3 used the 86 generic behaviors identified in Study 1 and the fourth dimension of the job element method (is trouble likely) to fish out the test misuse importance of each of the 86 items as they relate to each test cluster. Essentially, sets of ratings were gathered across different types of tests to determine if different behaviors are more relevant to different types of assessments. For example, statistical knowledge and being well versed in the psychometric principles of assessments may be more relevant to individual intelligence tests than to ability and preference assessments.

Study 3 successfully demonstrated that certain behaviors or knowledge factors are more important for the competent use of some tests than for others. The seven factors that must be involved in test use include: comprehensive assessment, proper test use, psychometric knowledge, maintaining the integrity of test results, accuracy of scoring, appropriate use of norms, and interpretive feedback. Through factor and cluster analyses, these seven factors were weighted relative to the eight types of assessments listed above. Assessments were than linked to “necessary skills” that approximately loaded at .80 or greater. Thus, the work group was able to state with some certainty that specific skills are essential for the use of a specific test. These findings represent a shift in test use emphasis. Previously, generic guidelines were proposed for all tests, but the work group revealed that varied levels of knowledge are more relevant to different types of tests.

Based on their findings, the TUQWoG developed an empirically derived test purchaser form. Borrowing from the knowledge above, test purchasers are required to possess the “necessary skills” that are empirically linked to assessments in a given cluster. The onus is still on test publishers to insure that only appropriately trained users have access to given tests, but it is clear that test publishers are armed with a better set of guidelines. Regardless of how thorough publishers are, ultimately the competent use of assessments is tied to the individual skills and competencies of clinicians.
The Task Force on Test User Qualifications (TFTUQ)

The American Psychological Association established the Task Force on Test User Qualifications (TFTUQ) in 1996. The goal of the task force was to develop guidelines that inform test users and the general public of the qualifications that the APA considers important for the competent and responsible use of psychological tests. The goal represents a significant shift in ideology of the APA. Previously, the emphasis was on test publishers and the appropriate distribution of testing materials to qualified individuals (Eyde et al., 1993; Tyler, 1986). The creation of TFTUQ was in response to the idea that test users, even users that are identified as qualified, may not possess the knowledge and skills that the APA considers desirable for maximum test use (APA, 2000). Years prior to this admission, several researchers (Aiken et al., 1990; Eyde et al., 1993; Tyler, 1986) argued that the majority of the problems associated with test misuse were related to the lack of competence of individual users and not with the ease with which publishers distribute tests. The TFTUQ and ultimately the APA’s mission was to identify test user knowledge gaps and develop guidelines that would encourage training programs to make curricular changes. These changes would provide future test users with a strong background in measurement theory and psychometrics along with improved skill in the administration, interpretation, and communication of test results. Additionally, improved guidelines might encourage individuals to obtain additional education and information to improve their use of psychological tests.

The TFTUQ defines the phrase test user qualifications as “the combination of knowledge, skills, abilities, training, experience, and where appropriate, credentials that the APA considers maximum for psychological test use (APA, 2000, pg. 6).” In an effort to improve test user qualifications, the TFTUQ outlined a set of guidelines that the APA would adopt. The APA’s purpose in developing these guidelines was to inform test users, training programs, regulatory and credentialing bodies, and the public about competencies the APA considers important for good test use. Two types of test user qualifications are outlined by the TFTUQ: generic qualifications that serve as a basis for most typical uses of tests and specific qualifications.

**Generic Qualifications**

As the name implies, generic qualifications describe general competencies linked to the
general use of tests across all settings. These qualifications are essential because they identify minimum qualifications that all users should posses. Generic qualifications or core knowledge and skills are deemed essential for the appropriate use of all tests (Turner et al., 2001). Generic qualifications include psychometric and measurement knowledge, selection of appropriate tests, test administration procedures, ethnic, racial, cultural, gender, age, and linguistic variables, and testing individuals with disabilities. Examples of psychometric and measurement knowledge include knowledge about descriptive statistics, scales and scores, reliability, and validity. Examples of test selection knowledge include understanding the intended use of a given test, scoring procedures, and possible test biases associated with a given test. Examples of test administration procedures include knowledge about the legal rights of test takers, standard administration procedures, and standard scoring procedures. Examples of ethnic, racial, cultural, gender, age, and linguistic variables include knowledge about construct equivalence, test bias, and procedures for examining between-group differences in test performance. Examples of testing individuals with disabilities include knowledge about legal issues, test accommodation, and the interpretation of test results to individuals with disabilities.

**Specific Qualifications**

Specific qualifications are those which describe specific competencies linked to the maximum use of tests in particular settings for specific purposes. The contexts in which tests are used include the setting and the reason the testing is being accomplished. Individual test user qualifications vary across and within settings (Guion, 1998; Turner et al., 2001). Regardless of the setting, according to the APA (2000), psychological tests are typically used to classify, describe, predict, track, and assist in intervention planning. Classification is defined as using test results to assist in placing clients in a specific taxonomic system or to aid in diagnosis. Description is defined as using test results to capture the strengths and weaknesses of an individual, group, or organization. Prediction is defined as using test results to predict future behavior based on previous data or a theoretical model of application. Intervention planning is defined as using test results to determine the appropriateness of different interventions or using test results to design a intervention model. Finally, tracking is defined as using test results to monitor psychological characteristics over time.

The TFTUQ delineated five different contexts in which generic or core skills and specific
skills are necessary in differentiating degrees. The five contexts include the employment context, educational context, career/vocational counseling context, health care context, and forensic context. While all five contexts warrant a thorough review, remaining consistent with the scope of this endeavor, this review will focus solely on the career/vocational counseling context and the specific qualifications outlined by the APA’s guidelines.

**Specific Qualifications: Career/Vocational Counseling Context**

Ultimately, psychological testing in the career counseling context is used to assist clients in decision making distress (Blustein & Spengler, 1995; Crites, 1976; Holland, 1994; Lucas & Epperson, 1988; Peterson, Sampson, and Reardon, 1996; Sampson et al., 1999; Super et al., 1979). This distress can be linked to mental health problems (Blustein & Spengler, 1995; Lucas & Epperson, 1988), faulty beliefs and negative attitudes (Krumboltz, 1979), problematic thought processes (Sampson et al., 1999), career immaturity (Crites, 1976), vocational identity (Holland, 1997) or simply indecision based on too little information (Crites, 1976; Super, 1979). The most common methods used to improve the amount of information clients possess about themselves are personality, ability, and interest inventories (Herr & Cramer, 1996; Holland, 1994; Zunker, 1994). Furthermore, one of the forefathers of career counseling, John Holland, developed an ingenious method of trait typing occupations and occupational personality types so clients might integrate knowledge about themselves with a “concrete knowledge” of a variety of occupations. Cognitive and mental health applications predominantly borrow from the counseling and health care fields that already exist, though a number of theorists have transformed cognitive theories from psychology to career specific contexts (Peterson, Sampson, & Reardon, 1996; Sampson et al., 1999).

According to the guidelines outlined by the APA (2000), test users in the career/vocational counseling context should be well versed in related areas of psychology, including adolescent and adult development, personality and psychopathology. Moreover, according to the APA, test users in this context must make every effort to be knowledgeable about types of work settings, work cultures and values, and the characteristics and requirements of types of jobs. Other systemic variables that can influence the person-environment fit include a client’s family system, gender, race, cultural background, physical ability, socioeconomic status, and psychological well being.
Career Context: Classification

Vocational classification is specifically defined as the act of coupling an individual’s career-related skills, abilities, and characteristics with the qualities and requirements of specific jobs or job categories (APA, 2000; Holland, 1994). Along with personality, cognitive, and interest inventories (Holland, 1994; Sampson et al., 1999), career counselors may use achievement and aptitude tests to determine a test taker’s special competencies (Kapes et al., 1994; Lowman, 1991). In the career context, classification as a specified qualification or competency is best defined as the ability of a counselor to integrate a variety measures into a meaningful tool of exploration and insight. Counselors must be able to effectively communicate with clients and illustrate how the results of a given test or series of tests outlines their person-environment fit as it relates to the client’s career options.

Career Context: Description

The ability of counselors to construct a holistic description of an individual’s personality and mental health is essential in the career/vocational counseling context (Gybers, Heppner, & Johnston, 1999). This specific competency is tied to a counselor’s ability to identify the maladaptive coexistence of mental health problems and career distress (Blustein & Spengler, 1995; Lucas & Epperson, 1988). The ability of a counselor to move from classification to description is adeptly captured by differentiating the levels of intervention between master’s level career advisors and doctoral level career counselors. Career advisors are best qualified for vocational classification whereas doctoral level career counselors are armed with the necessary experience and education to address severe personality disorders and other related psychological maladies that might impede career development. Thus, description as a qualification in the career context is best defined as the ability to assess clinically significant mental health problems and personality dysfunction as these impairments relate to career distress.

Career Context: Prediction

Prediction is arguably at the heart of the career context. Most often, career counselors use tests to capture stable, enduring traits that are tied to future work performance and satisfaction (Holland, 1994). However, career counselors must be cognizant of the developmental level of
test takers. Many theorists agree that individuals may lack the experience and knowledge that is associated with vocational crystallization (Blustein et al., 1994; Tinsley & Barrett, 1977). The APA advises, test users should modify predictions of future behavior with an understanding that test takers’ continued development may strongly influence their work behaviors. Prediction as a competency in the career context is best defined as the ability of counselors to recognize the traits and attributes of clients that are most relevant for each individual’s success and satisfaction in work, controlling for developmental and circumstantial influences that may influence their current test results.

**Career Context: Intervention Planning**

Career interventions range from brief interviews to extensive therapy. Regardless of the level of intervention, interventions quite often involve some mode of test administration and interpretation (Lowman, 1991; Tinsley & Bradley, 1986; Zunker, 1994). When test users are primarily concerned with improving a client’s self-knowledge through interest inventories, personality profiles, and value or belief inventories, test users should be skilled at promoting their client’s active engagement in the process (Tinsley & Bradley, 1986). When test users are primarily concerned with determining the factors associated with more serious complications like chronic unemployment and severe vocational distress, test users should be adept at coupling psychological assessments with career assessment tools and techniques (Fouad, 1994). Finally, when test users identify clients that lack career knowledge, are significantly indecisive, or who are vocationally immature, test users should be adept at selecting the appropriate assessments and intervention techniques to improve their client’s career situation (Crites, 1976; Holland, 1994; Super, 1979). For example, a client may show very little differentiation after completing a Holland typology test but also have very little knowledge about the world of work and the jobs that are available. A test user might be inclined to instruct the client to complete an internship, volunteer at a placement that fits the client, or take a part-time job and to return after three months for further work with the counselor. Intervention planning as a competency is best defined as the ability of a counselor to select and identify the appropriate assessment tools and intervention strategies necessary to provide the client with information that is understandable and meaningful to their needs at the time of their intervention.
Career Context: Tracking

Tracking as a contextual qualifier is best described as testing that allows counselors to compare patterns of growth within groups and/or individuals. As a competency, a counselor should be adept at identifying the changes that occur in a client both immediately and in the future. For example, the client from the section above may return in two months after holding a part-time job and still show similar undifferentiated test scores using the Holland typology test. However, the client appears to be exhibiting less indecision. A counselor might use another instrument like the Career Thoughts Inventory (Sampson et al., 1996) and determine that the client exhibits no career commitment anxiety, no career decision making confusion, and no external conflict from family or friends. From here, the counselor might conclude that the client is still developmentally and professionally quite young and that his undifferentiated scores are a healthy part of being nineteen years old. Moreover, his scores on the CTI are low enough to allow the client to make decisions when his level of experience, developmental growth, and education increases.

Regarding tracking, counselors must be knowledgeable about the psychometric and context-related implications of assessing career development over time. Again, tied to the example above, a counselor should immediately be concerned with the test-retest reliability of the instruments used above. Still more, the test user should be aware of any psychometric relationship between the SDS and CTI as well as the overall validity and reliability of the measures used above and any subsequent measures that might be used to track a client’s progress in the future. Tracking as a competency in the career context is best defined as the ability of a counselor to select, administer, and interpret initial and follow-up test results in a manner consistent with the client’s problems and the psychometric principles that govern the test or tests selected.

Qualifications: Summary

In tracking the progression of events to date, in 1950, the APA set out to establish guidelines that would assist test publishers in the distribution of psychological assessments. Thirty years later, the APA acknowledged that the system they had developed in 1950 was outdated. Testing had become much larger than any of the APA’s forefathers could have
imagined. In 1984, the APA merged with the AERA and NCME and developed a Joint Committee of Testing Practices. This was a merger of several associations committed to establishing more coherent guidelines for test use. From the JCTP, a work group was created to develop a data-based approach to promoting good test use. Again, the emphasis of this workgroup (TUQWoG) was to develop guidelines that would assist test publishers in the dissemination of testing materials.

Shortly after the work group completed their task, the APA acknowledged that test misuse was not entirely tied to publishers and their respective distribution practices. In fact, they admitted that is was likely, the majority of test misuse was tied to the misguided or misinformed practices of test users (APA, 2000; Turner et al., 2001). In 1996, the APA assembled the Task Force on Test User Qualifications (TFTUQ). The TFTUQ was established because the APA believed there was evidence that “some current users of psychological tests may not possess the knowledge and skill that the APA considers desirable for maximum test use” (APA, 2000, page 6). The task force borrowed from the work of the TUQWoG and developed two global sets of test user qualifications: generic and specific.

Generic qualifications are basic competencies, knowledge, or skills that all test users should possess. These competencies include basic knowledge about statistics, measurement, and diverse populations. Specific qualifications are best described as five basic competencies that are applicable across different practice contexts. These qualifications vary according to context and test users should possess specific knowledge in specific areas when using assessments in a given context. The five competencies include classification, description, prediction, intervention planning, and tracking. These specific competencies include specific guidelines about when a test is used, why a test is selected, whom a test is administered to, how results are interpreted, and what is shared with clients. The five contexts that these competencies apply to in varying degrees include the employment context, educational context, career/vocational counseling context, health care context, and forensic context.

For this endeavor, the career/vocational context and the competencies relative to this domain were explored. The career/vocational counseling contextual guidelines emphasize the importance of properly selecting instruments, differentiating which assessments are appropriate for which clients given their developmental level, and differentiating between the most important and developmentally appropriate forms of information you share with clients. Given this, it is
essential that career/vocational counselors know why they select the tests they do, when to administer tests, what information they share with clients, and how they integrate information gained from tests in a meaningful and competent manner.

**Rationale for the Development of a New Model of Competency Assessment**

One might imagine that a historical overview of the standards of practice, ethical conduct of clinical assessment, and the history of test use practices would be much larger in scope than was presented above. The truth is that research in this area has been minimal until recent years. While this may seem illogical, it makes sense after one attempts to operationalize concepts like abstract thinking and the assessment of abstract processing in trainee use of tests. Content domains are much easier to assess and make up the majority of the competency testing in the field of psychology. Several of the authors referenced in the above sections have expressed a great deal of concern about counselor trainee competency. They argue that counselor training programs are failing to adequately prepare future clinicians for clinical assessment (APA, 2000; Baldo et al., 1997; Custer, 1994; Eyde et al., 1988; Frame & Stevens-Smith, 1995; Knoff & Prout, 1985; Lumadee & Duffey, 1999; Meyer, 1980; Robertson & Eyde, 1986; Zimpfer et al., 1997). Moreover, nearly all of the authors cited above reported being surprised by the paucity of literature in this area. Based on the summary above, guidelines for test distribution and competency guidelines for test use have been comprehensively delineated. However, to date there are no tools available to determine if the guidelines put forth are being met. Clearly, the next step is the development of a competency assessment model that can measure the abstract level in which trainees are operating when using tests, a model designed to assess the specific competencies outlined by the American Psychological Association (2000). As stated earlier in this review, any set of guidelines, qualifications, or competencies are only as effective as the measures used to verify their proper use. Before describing the steps involved in constructing this model, a review of the role that cognition plays in counselor development is in order.

**Competency and Counselor Development**

Competency researchers argue that competencies are observable, measurable, and containable (Hatcher & Lassiter, 2004; Kaslow, 2004; Kaslow, Borden, Collins, Forrest, Illeder-Kaye, & Nelson, 2004; Krishnamurthy, VandeCreek, Kaslow, Tazeau, Milville, & Kerns, 2004;
Skovholt & Ronnestad, 1992; Stratford, 1994). Additionally, all of these researchers agree that competency grows in a developmental, stage-like manner. Across all practice oriented fields, the educational models used to train budding practitioners reflects an inherent understanding that cultivating competency is achieved by subscribing to a developmental model of training. For example, medical student education uses a problem-based learning model (Evensen & Hmelo, 2000) that augments classroom learning with real world vignettes. As these students show the ability to link small constellations of symptoms to accurate diagnoses, students are presented with increasingly complicated vignettes. Kaslow (2004) notes that it is imperative that competency-based education and training be developmentally informed.

While several theorists have introduced varying concepts of competency measurement (Dreyfus & Dreyfus, 1986; Skovholt & Ronnestad, 1992; Stoltenberg, 2005; Stoltenberg, Solomon, & Ogden, 1986), all seem to agree that competence is achieved through a marriage of cognitive analysis and stage-like developmental processes. Competency development in the use of assessments is best described as a stage-like process that involves the introduction of a given assessment to a trainee, followed by preliminary action steps of trainees orienting themselves to assessments, followed by action steps and analyses associated with the repeated use of assessments. Competency grows when action steps and the analytic processes necessary to properly use an assessment become internalized and increasingly automatic (Hatcher & Lassiter, 2004).

As noted above, counselor developmental processes can be described in many ways (Dreyfus & Dreyfus, 1986; Skovholt & Ronnestad, 1992; Stoltenberg, 2005; Stoltenberg, Solomon, & Ogden, 1986). Dreyfus and Dreyfus (1986) define five stages of counselor development. These stages are novice, advanced beginner, competent, proficient, and expert. For Dreyfus & Dreyfus, as trainees become more familiar with the analytic processes and actions of given tasks, performance becomes more integrated, flexible, efficient, and masterful. Stage to stage progression hinges on a weighted mastery of each stage. Weighted or staggered mastery is a form of stage progression that does not require individuals to complete one stage to move to the next stage. Weighted mastery occurs when individuals learn and integrate enough of the information from an early stage to progress to the next stage. While this individual will benefit from building on the preceding stage, he or she is capable of performing at a competency level of the higher stage.
Piaget

For Piaget, cognitive development progresses from the pre-logical to logical through four increasingly complex levels (Flavell, 1963; Flavell & Markman, 1983; Piaget, 1970; Piaget, 1977). Each developmental level begins with a structure and modus operandi intrinsic to a respective level and ends with the restructuring and integration of previous stages. The process of equilibration governs the acquisition of knowledge as an individual strives for the appropriate intellectual balance between stability (i.e., assimilation) and change (i.e., accommodation). As a result of interactions between existing schemas and the demands of the environment knowledge structures become more comprehensive, complex, and adaptive at each level of higher understanding.

The present research is focused on the logical thinking and new capacities associated with the advanced stages of Piaget, the concrete operational stage and the formal operational stage. Operational thinking begins the extension of thinking from the actual to the potential. Concrete operations are limited to physical reality and what an individual can manipulate, whereas formal operations are oriented towards ideas and the abstract. The formal thinker can think symbolically and scientifically, whereas abstract concepts resist explanation for the concrete thinker. While the stages are overlapping and continuous, the products of logical thought are characteristically distinct at each level. The rate at which development progresses varies, with some individuals never developing the mental abilities characteristic of formal operations regardless of the amount of experience with the learning material. Intellectual development is always in transition and many individuals respond in ways that may be characteristic of both concrete and formal operations. Since the stages are not discrete, transitional thinking must occur upon entry into formal operational thinking. During this transitional stage, concrete operations continue with the results being integrated into a more comprehensive system.

The concrete stage is characterized by face value interpretations and empty or glib elaboration completely tied to the instrument and constrained by reality or context. Transitional processing is marked by an ability to integrate the assessment with the client’s presenting problem. This is best described as a construct to counseling competence in the identification of a problem and the use of the assessment or decontextualization. Formal processing is characterized
by a recontextualization of construct dimensions. This requires a higher ordered abstraction when using assessment instruments and when incorporating the results of any assessment into a given clients reality. In formal processing, advisors integrate the instrument using feedback, dialogue, the results of the instrument, and the understandings that arise from each of these perspectives. The levels of processing (concrete, transitional, and formal) and accompanied criteria represent the operational components of this model.

**Developing a Measure of Competency**

In an attempt to further examine the factors associated with the conceptual misunderstandings of career advisors regarding the use of the CTI, an interview format that assessed the level of conceptualization at each step of the test use process was generated. Borrowing from the standards outlined by the APA, the conceptual progression of test use included: when do trainees decide on the use of the CTI, how do trainees conceptualize the meaning of the results of the CTI, and how do trainees interpret and share the results of the CTI in a meaningful manner when working with clients.

When reviewing the responses from trainees regarding the use of the CTI, the three primary developers of the instrument recognized distinct patterns of abstract processing between and within participants. One of the principle developers of the CTI compared the three levels of responses from trainees to that of Jean Piaget’s cognitive developmental model. That is, several trainees appeared to be operating at a very concrete level of thinking while others appeared to be operating at a more formal level of thinking as they described their use of the CTI. Even more interesting, the majority of the participants appeared to be operating at a transitional level between concrete and formal. For example, trainees routinely exhibited signs of formal operational thinking regarding the selection of the CTI but their rationale for selection failed to meet a criterion consistent with formal operative processing. Based on content analyses only, the participants appeared to be operating at a very high level (e.g., accurate generalizations, parroting jargon). However, an analysis of their more formal processing or their rationale for use revealed selection and decision criteria below what the APA might deem as fit. Again, borrowing from Piaget, their level of abstraction appeared firmly rooted in concrete operational thinking. In an effort to empirically quantify trainee level of cognitive operation, a matrix was created borrowing from Piaget’s cognitive theory and the competencies outlined by the APA above.
Piaget’s model of cognitive development was used to delineate three distinct levels of performance competency (Flavell, 1963; Flavell & Markman, 1983; Piaget, 1970; Piaget, 1977).

**Summary of the Matrix Formation and the Next Step**

The theoretical model known as the Piagetian Matrix of Test User Competence (PMTUC) and the instrument born from this theory (AMAC-CTI) were developed from a pilot study at FSU that combined the career/vocational counseling guidelines put forth by the APA, the expertise of three faculty members at FSU, and the levels of cognitive processing delineated by Jean Piaget in his famous work on cognitive development. This AMAC-CTI was developed to determine if an instrument could be created that successfully identifies the overall level of test user competency. Moreover, the items in the AMAC-CTI, while not intended to be subscales at this time, can offer insight as the cognitive criteria that counselors in training use to select, administer, and interpret psychological tests. The AMAC-CTI was developed to assess the process mechanics, not the knowledge or content domain, of test users or counseling trainees. The development of the AMAC-CTI follows the guidelines recently completed by the APA (2000). The next logical step is to use the PMTUC theoretical model to test an instrument (AMAC-CTI) that allows us to apply our theory. The AMAC-CTI, originally comprised of 10 items, now contains 6 questions designed to measure competency development. The current purpose of the AMAC-CTI is to provide one global competency rating of test use. Trainee responses to the six questions can be used to guide remediation efforts but the current intent of this researcher is to create an instrument that provides one score.

**Building the Instrument**

In the fall of 1998, the Center for the Study of Technology in Career Counseling and Development at a large southeastern university (FSU) initiated a pilot study to determine the competency exhibited by career advisors in the use of assessment instruments. The Career Thoughts Inventory (CTI) was selected as the psychological assessment in this study. The CTI was selected because the originators of this instrument were three faculty members at FSU. Just as the work group and task force above sought experts, the pilot study FSU embarked upon required a panel of experts to assess career advisor performance. The originators of the CTI appeared to be good candidates as experts. The CTI is an assessment tool that can be used by
career counselors to develop interventions designed to identify, challenge, and alter
dysfunctional cognitive processes.

Borrowing from the revised standards outlined by the APA (2000), a series of questions
were formulated by the experts above. The questions read as follows:

• From your understanding, what is the purpose of the CTI?
• From your understanding, what constructs does the CTI measure?
• On what basis do you assign the CTI?
• On what basis do you not assign the CTI?
• Give an example(s) when you use the CTI.
• Give an example when you do not use the CTI.
• What steps do you take when interpreting the results of the CTI with clients?
• What information do you share?
• What steps do you take when developing an Individual Learning Plan with CTI data?
• What interventions do you prescribe that are linked to the CTI?
• Give an example of an intervention you have used.

The results of the pilot study revealed that the career advisors had difficulty using the
assessment. Specifically, they used irrelevant criteria as a basis of administration and failed to
demonstrate adequate competency in the interpretation of results to clients. These results violate
competencies outlined by the American Counseling Association (1995) and the American
Psychological Association (2000). These questions were designed to cull out abstract processing
related to the generic and specific qualifications listed in the sections above. From the generic
domain, the questions above determine the level of competency of the counselor trainees in test
selection (e.g. purpose of the CTI, constructs, rationale for use) and test administration (e.g. steps
involved, information shared). From the specific qualification domain and more specifically, the
career context of intervention planning, the question regarding ILP development and the question
regarding interventions prescribed address the competency in which trainees use the CTI.

Process Mapping: Steps toward Exemplary Test Use

Based on these findings, determining the nature of the misuse of the CTI was undertaken
by outlining the conceptual progression associated with the use of the CTI. Using subject
responses to the questions above coupled with the competencies involved in the selection, administration, and interpretation of assessment instruments, three expert faculty members and one graduate student worked to develop a process map that outlines exemplary test use. A graphical representation of this map is presented in Figure 1.

![Diagram](image-url)

**Figure 1. Competency Flowchart**

Again, the FSU Center of Technology in Career Studies decided the next logical step after promulgating guidelines was to develop a competency assessment model based on the task
force’s guidelines, a model that captures the level in which counselor trainees are emulating the standards put forth by the APA. After deciding on the CTI, the guidelines created by the task force and promulgated by the APA were incorporated into a design that asked trainees when, why, and how they use the CTI. Determining how to assess the level in which graduate trainees met the guidelines created by the task force was no easy task. This involved quantifying the clinical rationales offered by budding counselors.

It was the intent of the author of this paper to create A Measure of Assessor Competence. The theoretical model was based on a series of questions formulated by a thorough review of the APA test use standards, subject responses to these questions from a 1998 pilot study, the collective expertise of 3 clinical trainers, and Jean Piaget’s model of cognitive development. A matrix was created by developing a criterion scale in which ratings of the trainee responses were delineated into one of three levels of cognitive articulation (Concrete, Transitional, Formal). The levels are described in detail in Appendix A.

The Next Step: Validating the Measure of Assessor Competence

The reliability coefficients reported in the field tests were very encouraging and represent another step toward building a valid assessment tool. As Yu (2005) stated, “reliability is a necessary but not sufficient condition for validity. If the needle of the scale is five pounds away from zero, any person using this scale will over-report their weight by five pounds. Is the measurement consistent? Yes, but it is consistently wrong. (p. 112)” The next step is to conduct another series of reliability and validity assessments. Chapter 3 outlines 5 studies that will be conducted to further validate the AMAC-CTI.

The first study was a content validation study that involves five expert raters. Experts judged the importance of each item on the AMAC-CTI. The second study examined the agreement among raters and the internal consistency among items. The third study was a measurement of task difficulty. The fourth study, concurrent validity, examined the level of agreement between AMAC-CTI scores and supervisor ratings of student participants. Supervisors used the same scoring template as AMAC-CTI raters. The fifth study examined the relationships between education and experience and performance on the AMAC-CTI.
CHAPTER 3
METHODOLOGY

The following methodology will describe studies of content validity, reliability, item
difficulty, and convergent validity. Convergent validity was assessed by comparing AMAC-CTI
domain and global scores to experience and education. Convergent validity was assessed by
comparing supervisor ratings of student counselors across the six domains of the AMAC-CTI
with respective student AMAC-CTI scores. Finally, inter-rater and inter-item reliability analyses
were conducted.

Study 1 – Content Validity

Purpose

The purpose of this study was to establish content validity for a measure of assessor
competence, specifically, the AMAC-CTI (see Appendix C). Content validity is defined as the
degree to which test content is tied to the instructional domain it intends to measure. For this
study, experts examined and rated the relevance of the 6 domains of the AMAC-CTI. The
AMAC-CTI was designed to assess the level of competence counselors and/or counselor trainees
exhibit in the use of psychological assessments. The assessment selected for this study was the
Career Thoughts Inventory (CTI). The AMAC-CTI can ultimately be used as a training and
remediation tool for professionals and counselors in training. The research question tied to this
study is, to what extent are the items that comprise the AMAC-CTI total score of competence
important indicators of competency.

Participants

Five experts served as raters for this study. One rater was a Master’s level clinician with
10 years of experience with the use of the CTI. One rater was the Director of Career Services
Center at a large southeastern university. Three of the five raters were faculty members at a large
southeastern university. Three of the five raters are also the creators of the Career Thoughts
Inventory.

Instrumentation

Career Thoughts Inventory (CTI): The Career Thoughts Inventory is a 48-item theory
based measure of readiness for career problem solving and decision-making. The CTI is used to identify dysfunctional thinking and subsequently develop interventions designed to alter maladaptive cognitive processes of high school students, college students, and adults. The internal consistency coefficients of the CTI overall score range from 0.93 to 0.97 and test-retest reliability is reported at 0.77. The psychometric properties of the CTI are well validated (content, convergent, criterion-related, and construct).

A Measure of Assessor Competence (AMAC-CTI): The instrument, A Measure of Assessor Competence was developed by the authors of the Career Thoughts Inventory. As part of a 1998 pilot study, nine interview questions were developed based on APA’s guidelines of test use. Subject responses in the pilot study were based on the supposed use or non-use of the Career Thoughts Inventory. The primary authors of the CTI and one graduate student characterized subject responses in one of three ways based on Piaget’s theory of cognitive development. If participants responded to a question in a vague and contextually limited manner, the response was judged to be “concrete”. A “transitional” response was typically less vague, not bound by general rules of use, and showed signs of abstract processing that involved client characteristics as well as the constructs outlined by the instrument. A “formal” response was typically clear, governed primarily by client needs, and linked to diagnostic and prescriptive qualities of instrument. The PMTUC theoretical model in which responses to the AMAC-CTI instrument were judged is completely delineated previously in Table 1.

The three levels outlined above were quantified so that a rating system could be used with future counselor trainee populations. Scores of 1, 2 or 3 were designated as concrete responses. A score of 3 indicated that the response was approaching a transitional level of explanation but failed to exhibit enough of the qualities of a transitional response to rate it so. A score of 1 indicated the response was completely tied to the instrument and is best characterized as a response filled with parroted jargon. Scores of 4, 5, and 6 indicated the response was at the transitional level, and scores of 7, 8, and 9 indicated that the response was at a formal level of operation. The same rules applied where a score of 6 represented a more competent score than a 4 within the transitional level and a score of 9 was better than a score of 7 in the formal domain. Reliability for this revised design was 0.89 (Etheridge et al, 2004). Moreover, when inter-rater agreement was reduced to selecting one of the three levels (concrete, transitional, and formal), agreement was nearly perfect at .96 (Etheridge et al, 2004). An inter-item analysis was
performed to assess the AMAC-CTI reliability. Based on this measure of internal consistency, two items were removed. The theoretical basis for inclusion was not compromised because the omitted items were closely related to two other items in the AMAC-CTI.

Expert Content Rating Form: The Expert Content Rating Form was developed for experts in the field of career counseling and/or career counseling theory. See Appendix H for the Expert Content Rating Form. Designated experts used this rating form to rate the content of the AMAC-CTI. The form contains an introduction section that contains an explanation of the purpose of the rating. After the introduction, a rating key was provided with the definition of ratings between 1 and 4. A rating of 1 indicated that the rater believes this item is of no importance. A rating of 2 indicated that the rater believes the item is of some importance. A rating of 3 indicated that the rater believes the item is important but not critical. A rating of 4 indicated that the rater believes the information supplied by the item is critical to the AMAC-CTI process.

Procedure

Five experts in the field of career counseling, including the authors of the assessment instrument used for this study, were asked to participate in this study. The purpose of this study was explained to each expert. Upon agreeing, each expert was given an Expert Content Rating Form (Appendix H) and a copy of the AMAC-CTI scoring matrix (Appendix C). Each expert was asked to judge the necessity or value of each item using the content rating form. The experts were also asked to rate the importance of each item as an indicator of assessor competence.

Statistical Procedures

Expert ratings of the content of the AMAC-CTI were examined using several techniques. The distribution pattern of all the ratings was examined to determine normality. Pearson product correlations were used to determine rater-to-rater and rater-to-group agreement. Rater bias was examined by comparing individual ratings to group mean ratings. Researchers planned to examine ratings that exceeded 2 standard deviations by interviewing raters regarding their rationale. Follow-up was not necessary.

Study 2 – Reliability

Purpose

The purpose of this study was to establish the reliability of the Measure of Assessor
Competence (AMAC-CTI). Reliability alludes to the precision of scores derived from the AMAC. Inter-rater reliability and internal consistency were assessed. Inter-rater reliability is defined as the degree to which the measuring instrument yields similar results at the same time with more than one assessor. Internal consistency is defined as an estimation of reliability based on the correlation among the variables comprising a set. The AMAC-CTI was designed to assess the level of competence counselors and/or counselor trainees exhibit in the use of psychological assessments. The AMAC consists of six variables and one overall score. There are three research questions associated with this study. These questions are: 1) To what extent was there agreement among raters with respect to task performance; 2) To what extent are the items that comprise the global AMAC-CTI measurement internally consistent; and 3) To what extent was the trait measured by the AMAC-CTI uni-dimensional?

Participants

Thirty-three total individuals participated in the survey portion of the study, eighteen of which were graduate students. Fourteen of the participants were doctoral level clinicians, one was a master’s level clinician, sixteen were doctoral students, and two were master’s students. The two master’s students were in their second year of study. Twelve of the doctoral students were post-master’s degree, and the others were doctoral students who had not completed a master’s degree prior to entering their doctoral program. With respect to education specific to assessment and test theory, the participants reported having had a mean of five courses in these areas, with a median of four and a range from one to fifteen. Nearly a third of the sample had four courses in assessment and/or test theory. Twelve percent of the survey sample were black, six percent were Hispanic, and eighty-two percent were white. Seventy-six percent of the survey sample were female. Twenty-four percent of the survey sample were male. Five expert raters participated in the rating task of this study. This task involved judging the responses of the thirty-three survey responses using the AMAC-CTI criteria. One rater is a master’s level clinician with 10 years of experience with the use of the CTI. One rater is the Director of Career Services Center at a large southeastern university. Three of the five raters serve as faculty members at a large southeastern university. Three of the five raters are the creators of the Career Thoughts Inventory. Some of the participant characteristics are presented in Table 1.
Table 1

*Participant Characteristics*

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Standard Deviation</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Education – Years Graduate Study</td>
<td>5.36</td>
<td>1.85</td>
<td>2 - 7</td>
</tr>
<tr>
<td>Courses in Assessment</td>
<td>5.00</td>
<td>2.77</td>
<td>1 - 15</td>
</tr>
<tr>
<td>CTI Administered</td>
<td>92.18</td>
<td>151.5</td>
<td>2 - 750</td>
</tr>
<tr>
<td>CTI Review with Supervisor</td>
<td>12.48</td>
<td>14.63</td>
<td>0 – 75</td>
</tr>
</tbody>
</table>

Graduate students enrolled in a Combined Counseling Psychology and School Psychology doctoral program or the Mental Health Counseling masters program at a large southeastern university were approached about volunteering for a study of trainee competency using assessments. Doctoral level and masters level professionals were asked to participate in this study by the principal investigator and the student conducting this research. Students and professionals were identified by the training director at the aforementioned southeastern university. The training director verified that all the participants selected to participate in this study have been trained in the use of the Career Thoughts Inventory. Only those professionals who have received CTI training were eligible for participation in this study. Potential participants were approached via face-to-face request, telephone request, or via e-mail.

**Procedure**

Graduate students enrolled in a Combined Counseling Psychology and School Psychology doctoral program or the Mental Health Counseling masters program at a large southeastern university were approached about volunteering for a study of trainee competency using assessments. Doctoral level and masters level professionals were asked to participate in this study by the principal investigator and the student conducting this research. Students and professionals were identified by the training director at the aforementioned southeastern university. The training director verified that all the participants selected to participate in this study have been trained in the use of the Career Thoughts Inventory. Only those professionals
who have received CTI training were eligible for participation in this study. Potential participants were approached via face-to-face request, telephone request, or via e-mail.

Once participants agreed to participate, participants were given a web address for the survey. The survey included an electronic consent, a brief questionnaire, and the six questions that represent the six items on the AMAC-CTI. The brief questionnaire contained questions about educational level, experience with the CTI, and attitude about the usefulness of the CTI. Participants contacted via e-mail were given the Internet address of www.surveymonkey.com for the survey upon agreeing to participate. Participants initially contacted by phone or face-to-face were asked to provide an e-mail address so that the web address for the survey could be sent to them via e-mail. Participants were informed that their name would be entered into a lottery for a dinner for two at any Bennigan’s restaurant in the United States. Students were informed that they could terminate their participation at any time without penalty.

The Internet address www.surveymonkey.com provides researchers with a web based survey instrument. Collectively, the survey process took approximately thirty minutes to complete. The brief questionnaire took approximately two minutes to complete. Upon completion of the brief questionnaire, participants were asked to select the complete questionnaire box on the electronic form (or “e-form”). This shifted the screen to the electronic interview (or “e-interview”). On the e-interview form, participants were reminded that their participation was voluntary and that they may end the process at any time without penalty. Participants were informed they could take as long as they needed to answer each question. Participants were informed to avoid placing any identifying information on their response sheets. Once the subject completed the six questions, participants selected the finish option. The e-survey prompted participants to review their responses. Participants were allowed to change their typed responses, and/or select the finished interview option on the form. Once participants selected the finish interview box, they were thanked for their participation.

**Instrumentation**

The Career Thoughts Inventory (CTI) and the AMAC-CTI were used for this study. These instruments are defined in Study 1 starting on page 30.

**Statistical Procedures**

The three research questions are: 1) To what extent was there agreement among raters with respect to task performance; 2) To what extent are the items that comprise the global
AMAC-CTI measurement internally consistent; and 3) To what extent was the trait measured by the AMAC-CTI uni-dimensional?

To answer question one, correlation analyses were conducted to examine the magnitude of agreement among raters and an average inter-rater agreement were performed. The overall agreement across all raters for the global of measurement of competence was calculated followed by inter-rater intra-class correlations between all raters for overall score and each item on the AMAC-CTI.

To examine the internal consistency of the items of the AMAC-CTI, an overall Cronbach-Alpha coefficient was computed followed by correlations that examine the quantitative relationship between each AMAC-CTI item and overall AMAC-CTI scores.

To examine the final research question regarding uni-dimensionality, an exploratory factor analysis was performed to ascertain the internal structure of the AMAC-CTI. The specification of this analysis was principal components with varimax rotation.

**Study 3 – Item Difficulty**

**Purpose**

The purpose of this study was to determine the relative difficulty of responding to the six items on the AMAC-CTI. The AMAC-CTI was designed to assess the level of competence counselors and/or counselor trainees exhibit in the use of psychological assessments. To assess the level of difficulty associated with the performance task within this research, three research questions were poised regarding task difficulty. These questions were: 1. To what extent was the task challenging to test takers? 2. To what extent was there variability in the performances of test takers? and 3. To what extent was there evidence of lack of effortful processing or maximum performance?

**Participants**

Five expert raters and the subjects described in Study 2 participated in this study. Once respondents completed the survey, the five expert raters used the AMAC-CTI Respondent Rating Form to rate the responses of the thirty-three survey responses using the AMAC-CTI rating criteria outlined on the rating form. One rater is a master’s level clinician with 10 years of experience with the use of the CTI. One rater is the Director of Career Services Center at a large southeastern university. Three of the five raters serve as faculty members at a large southeastern
university. Three of the five raters are the creators of the Career Thoughts Inventory.

**Procedure**

Subjects were enlisted to participate in the same manner as outlined in Study 2.

**Instrumentation**

The instruments used for this study included the CTI and the AMAC-CTI. These instruments are defined in study 1 above.

**Statistical Procedures**

The three research questions within this study were: 1. To what extent was the task challenging to test takers? 2. To what extent was there variability in the performances of test takers? and 3. To what extent was there evidence of lack of effortful processing or maximum performance?

To examine research question one, descriptive statistics that include the mean, standard deviation, range, skewness, and kurtosis for each item were computed. To answer question two, the distribution of performance ratings was reported. To answer question three, the number of responses rated as zero were reported.

**Study 4 – Convergent Validity**

**Purpose**

The purpose of this study was to establish convergent validity for A Measure of Assessor Competence (AMAC-CTI). Convergent validity is defined as the degree to which a measure yields a result consistent with the results of another measure of the same construct. Concurrent validity was not selected because to demonstrate concurrent validity, the measure being used to validate a new instrument must be a well-validated, firmly established instrument (Smith & Glass, 1987). The AMAC-CTI was designed to assess the level of competence counselors and/or counselor trainees exhibit in the use of psychological assessments. The assessment selected for this study of assessor competence was the Career Thoughts Inventory (CTI). AMAC-CTI scores were correlated with supervisor ratings to determine the degree of convergence. The comparative ratings provided by supervisors in this study were based on direct observations of performance in the use of the CTI by the participants. These ratings were not products of a formal instrument confirmed by the rigors of psychometric validation. The research question that was addressed by this study is, to what extent was task performance on the AMAC-CTI correlated with supervisor
Participants

Participants in this study included the clinical supervisors of the students that agreed to participate in the survey portion of Study 2. The supervisors are employed by a large southeastern university career center. The supervisors included one doctoral level faculty member and one master’s level clinician with extensive experience with the use of the Career Thoughts Inventory. Thirty-three total individuals participated in the survey portion of the study, eighteen of which were graduate students. These eighteen students participated in this portion of the study. Sixteen of the eighteen were doctoral students and two were master’s students. The two master’s students were in their second year of study. Twelve of the doctoral students were post-master’s degree, and the others were doctoral students who had not completed a master’s degree prior to entering their doctoral program. With respect to education specific to assessment and test theory, the participants reported having had a mean of five courses in these areas, with a median of four and a range from one to fifteen. Nearly a third of the sample had four courses in assessment and/or test theory. One of the student participants was Hispanic and one was Asian. Four of the eighteen student participants were African-American. The remaining twelve student participants were White.

Instrumentation

The instruments for this study included the Supervisor Rating Form, the CTI, and the AMAC-CTI. The AMAC-CTI and CTI are defined in Study 1. The Supervisor Rating Form was based on the six domains of the AMAC-CTI. This form was developed for clinical supervisors. This rating form was used by supervisors to rate students on the six domains of the AMAC-CTI. See Appendix D for the Supervisor Rating Form.

Procedure

Student subjects who agreed to participate in the study were asked to identify their clinical supervisors, and those supervisors were asked to participate in the study as well. Students and supervisors were informed that the supervisors’ participation would involve providing a global competency rating for each student participant. During the student participant recruitment process, students who agreed to participate were asked to sign a consent form that would allow their supervisor to complete a rating form (Supervisor Rating Form – Appendix H). If agreeable, students were contacted in person by the primary researcher and were informed that
their supervisors would be asked to estimate their respective student’s competency in the use of the CTI based on the rating criteria designed by the researchers of this study. Students were informed that these estimations would be compared to scores derived from survey responses. They were also informed that their survey responses would not be shared with their supervisor. This consent was simply a request for permission to ask their supervisors to complete the Supervisor Rating Form. Finally, students were informed that they had the right to ask their supervisors about the ratings they provided.

Supervisors were informed of the aforementioned, and all supervisors agreed to talk with students if approached. The rating was based on the scoring criteria and global score format of the AMAC-CTI. Each subject had only one global rating from one supervisor. As noted above, the purpose of this study was thoroughly explained to all supervisors and students. The supervisors were given the AMAC-CTI scoring criteria matrix (Table 1) with the outline of each competency level via face to face hand delivery.

**Statistical Procedures**

The research question for this study is: 1) To what extent do competency ratings as measured by the AMAC-CTI correspond with supervisor ratings of competence using the same rating criteria. The supervisor ratings on the Supervisor Rating Form were identical to the rating scale on the AMAC-CTI. Since the average number of ratings per rater was small, an ANOVA to examine the variation among raters was not used. Instead, correlational analyses were used to measure the degree in which supervisor ratings of students and AMAC-CTI ratings agree regarding each student participant’s competence. To add, a visual analysis of frequency statistics was used to compare supervisor ratings to AMAC-CTI ratings. Finally, a paired sample t-test was performed to examine whether the difference between supervisor ratings and AMAC-CTI ratings differed significantly.

**Study 5 – Convergent Validity**

**Purpose**

The purpose of Study 5 was to further establish convergent validity for the instrument, A Measure of Assessor Competence for the Career Thoughts Inventory (AMAC-CTI). The relationship between AMAC-CTI scores (items and global) and subject experience and education level will be assessed. The AMAC-CTI was designed to assess the level of competence
counselors and/or counselor trainees exhibit in the use of the CTI. The AMAC-CTI can ultimately be used as a training and remediation tool for professionals and counselors in training. The research question being addressed by this study is, to what extent were scores on the AMAC-CTI associated with education level, the number of assessment courses completed, the number of Career Thoughts Inventories administered, and the number of Career Thoughts Inventories reviewed with supervisors.

**Participants**

The sample of interest included current master’s level and doctoral level graduate students from a large southeastern university and professionals in the field of psychology that currently use the CTI. The doctoral students were enrolled in the Combined Program in Counseling Psychology and School Psychology. The master’s students were enrolled in the Mental Health Counseling or Career Counseling program. The mental health professionals were employed at various client service facilities in the United States. The population of interest was students and professionals who have been trained in the use of the CTI.

**Instrumentation**

Education and Experience Brief Questionnaire: The brief questionnaire included four questions: What is your current education level? How many courses in assessment and test theory have you completed? Approximately, how many CTIs have you administered? Approximately, how many CTIs have you administered and reviewed with a supervisor?

Interview Questions: Six interview questions were developed using APA’s guidelines for test user qualifications (2001). The questions sought to assess the general and specific mental processes that participants use when selecting the Career Thoughts Inventory (CTI). The six questions are outlined in the previous section entitled “Building a Model of Competency Assessment”. The questions are a merger of the APA standards and the proposed use or non-use of the CTI.

The CTI and the AMAC-CTI were used in this study. These instruments are defined in Study 1.

**Procedures**

Subjects were enlisted to participate in the same manner as outlined in Study 2 above.

**Statistical Procedures**

The research questions for this study include: 1) To what extent was there agreement
between AMAC-CTI ratings and educational level?; 2) To what extent was there agreement between AMAC-CTI ratings and counselor experience in the use of the CTI; 3) To what extent was there agreement between AMAC-CTI ratings and the student coursework in assessment; and 4) To what extent was there agreement between AMAC-CTI ratings and the number of CTIs administered and reviewed with a supervisor.

Correlation analyses were performed to address all of these questions. Correlations between AMAC-CTI ratings and education, coursework, CTI administrations, and supervisor enhanced administrations were performed. The educational and experiential domains were correlated to global AMAC-CTI ratings and each item on the AMAC-CTI.

**Declaration of Statistical Significance**

For the purposes of this study, the alpha level to reject or affirm a hypothesis was set at 0.05. With an n=33, this level was set to be sensitive to relationships to reduce the threat of Type II errors, but also to reduce the threat of occurrence of random effects, i.e., Type I errors.
CHAPTER FOUR

RESULTS

Introduction

This research was designed to validate a measure of competency in the use of psychological assessments: A Measure of Assessor Competence (AMAC-CTI). The validation process consisted of five studies. The first study, content validation, involved five expert raters being asked to judge the relative merit of each of the items that comprise the AMAC-CTI. The second study examined three inquiry questions pertaining to reliability and factor analysis. These examinations included an investigation of inter-rater reliability and inter-item reliability. The third inquiry question was examined by an exploratory factor analysis. The third study was a measurement of task difficulty. Observational measurements of the mean performance of the experimental cohort were assessed. The fourth study was an investigation of the concurrent validity conducted by comparing AMAC-CTI scores to supervisor ratings of student participants. The fifth study examined the relationships between education and experience and performance on the AMAC-CTI.

The AMAC-CTI was designed to derive one global score or rating of competence in the use of psychological assessments. This one rating is based on a mean rating across six items. The six items represent questions about 1) a counselor’s understanding of the purpose of an instrument, 2) the constructs or domains that make-up an instrument, 3) the steps involved when administering a psychological assessment instrument, 4) the steps involved when interpreting the results of a given instrument, 5) the steps involved in developing a treatment plan based on the results of a psychological assessment instrument, and 6) the types of interventions prescribed based on the results of a psychological assessment. The tables that follow contain row and column headings that correspond with the AMAC-CTI items described above. The labels for the aforementioned in order of mention are Overall, Purpose, Construct, Use, Interpretation, ILP, and Intervene.

For the purposes of the construction of this measure, competency is theoretically conceptualized in Piagetian terms as Concrete, Transitional, or Formal. Concrete understanding is considered a contextually based understanding borrowing largely on the conceptual
memorization of a given psychological instrument’s properties. A rating of 1 represents a concrete rating by a rater for an item. Transitional operation is indicative of a counselor’s ability to partially decontextualize the properties of the instrument in client terms. This level of competence is indicated by a rating of 2 for a given item. Formal operation is indicative of a counselor’s ability to completely decontextualize the properties of a psychological assessment instrument in client terms. A rating of 3 indicates that the counselor being rated has responded to a question about purpose, constructs, use, interpretation, treatment plan development, or interventions prescribed in a highly professionally evolved manner. If a counselor’s response to one of the six items is completely void of meaningful information, a score of 0 is possible.

Since the overall rating of a given participant is the mean of six ratings, the single global score of competence derived from these ratings can be between 0 and 3.0, where a global rating of 0 to 1.4 indicates a level of concrete operation, 1.5 to 2.4 indicates transitional operation, and 2.5 to 3.0 indicates formal operation.

For the purposes of this study, as indicated in Chapter 3 of this dissertation, the psychological assessment instrument used to demonstrate competency of assessment was the Career Thoughts Inventory (CTI). When building a theoretical model of competency measurement and more specifically, a measure of competency in the use of psychological assessments, it was important to secure a psychometrically sound instrument and to secure the cooperation of professionals who are experts in the use of that instrument. There are no greater experts in the competent use of the Career Thoughts Inventory than the creators of the instrument, who were available and willing to participate in this research. The relationship between each of the items is listed in Table 2.

Table 2

Pearson Correlations

<table>
<thead>
<tr>
<th></th>
<th>Purpose</th>
<th>Constructs</th>
<th>Use</th>
<th>Interpret</th>
<th>ILP</th>
<th>Intervene</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall</td>
<td>.831(**)</td>
<td>.750(**)</td>
<td>.852(**)</td>
<td>.891(**)</td>
<td>.845(**)</td>
<td>.888(**)</td>
</tr>
<tr>
<td>Purpose</td>
<td>-</td>
<td>.733(**)</td>
<td>.622(**)</td>
<td>.674(**)</td>
<td>.581(**)</td>
<td>.676(**)</td>
</tr>
<tr>
<td>Constructs</td>
<td>-</td>
<td>-</td>
<td>.549(**)</td>
<td>.549(**)</td>
<td>.437(*)</td>
<td>.646(**)</td>
</tr>
<tr>
<td>Use</td>
<td>-</td>
<td>-</td>
<td>.757(**)</td>
<td>.693(**)</td>
<td>.656(**)</td>
<td></td>
</tr>
</tbody>
</table>
Table 2 - continued

<table>
<thead>
<tr>
<th>Purpose</th>
<th>Constructs</th>
<th>Use</th>
<th>Interpret</th>
<th>ILP</th>
<th>Intervene</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interpretation</td>
<td>-</td>
<td>.721**(*)</td>
<td>.767**(*)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ILP</td>
<td>-</td>
<td></td>
<td>.752**(*)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intervene</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-</td>
</tr>
</tbody>
</table>

**Correlation is significant at the 0.01 level (2-tailed). *Correlation is significant at the 0.05 level (2-tailed).**

**Study 1 – Content Validation**

The content validation of the AMAC-CTI was conducted to address the inquiry question: To what extent are the items from the AMAC-CTI perceived as important indicators of assessment competency by expert judges? Five experts were asked to rate the importance on each item in the AMAC-CTI on a four-point scale where a rating of 1 = not important to the measurement of competency, 2 = somewhat important, 3 = important, and 4 = critical to the measurement of competency.

Regarding the content of the AMAC-CTI, all five experts who completed these ratings agreed that all six items were important in determining counselor competency. Expert importance ratings and averages are presented in Table 3. There was no rating below 3.0 across all 30 possible ratings. The average agreement among raters was 75%. The mean inter-rater agreement was .72 with a standard deviation of .14. The range of agreement was .66 to 1.0. Borrowing from the assumption that an average rating of 3.5 or higher indicates that the experts agree that a given item is critical to the measurement of competency using the AMAC-CTI, only item 5, participant’s ability to describe treatment plan development based on the results of the CTI, was below the 3.5 cut-off. Understanding the purpose and the constructs of the CTI received the highest average endorsement possible of 4.0. Individual ratings were compared to mean ratings to explore rater bias. No ratings exceeded two standard deviations from the mean therefore no follow-up interviews with raters were necessary.
Table 3

**Expert Ratings of the Importance of AMAC-CTI Items**

<table>
<thead>
<tr>
<th>Item</th>
<th>Expert 1</th>
<th>Expert 2</th>
<th>Expert 3</th>
<th>Expert 4</th>
<th>Expert 5</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>Understands purpose of CTI</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4.0</td>
</tr>
<tr>
<td>Understands constructs</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4.0</td>
</tr>
<tr>
<td>Understands when to use</td>
<td>3</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>3.6</td>
</tr>
<tr>
<td>Interpretive Steps</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>3</td>
<td>3.8</td>
</tr>
<tr>
<td>ILP/Tx plan development</td>
<td>3</td>
<td>3</td>
<td>4</td>
<td>3</td>
<td>3</td>
<td>3.2</td>
</tr>
<tr>
<td>Specific Interventions</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>3</td>
<td>3</td>
<td>3.6</td>
</tr>
</tbody>
</table>

Table 3—continued

**Study 2 – Reliability**

This study addressed three inquiry questions: 1) To what extent was there agreement among raters with respect to task performance?; 2) To what extent are the items that comprise the global AMAC-CTI measurement internally consistent?; and 3) To what extent was the trait measured by the AMAC-CTI uni-dimensional?

**Reliability Question 1**

To what extent was there agreement among raters with respect to task performance?

The Cronbach-Alpha coefficient for total scores across all 5 raters for the global measurement of competence was .95. Specific intra-class correlations are presented in Table 4. Inter-rater correlations ranged from .57 to .91. All correlations were significant at the .01 level. The average inter-rater correlation was .60 with a standard deviation of .19. The general level of agreement among raters was consistent for the overall measure of competency.

Table 4

**Interrater Intra-class Correlations for the Overall Score**

<table>
<thead>
<tr>
<th></th>
<th>Expert 1</th>
<th>Expert 2</th>
<th>Expert 3</th>
<th>Expert 4</th>
<th>Expert 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Expert 1</td>
<td>-</td>
<td>.807(**)</td>
<td>.811(**)</td>
<td>.860(**)</td>
<td>.814(**)</td>
</tr>
</tbody>
</table>

45
The Cronbach-Alpha coefficient for total scores across raters for the Purpose item was .83. Inter-rater correlations ranged from .14 to .69. The intra-class correlations are presented in Table 5 for this item. 80% of the correlations were significant at the .01 level. 10% were significant at the .05 level and 10% were not significant. The general level of agreement for the Purpose item was moderate to high.

The Cronbach-Alpha coefficient for total scores across raters for the Constructs item was .89. Inter-rater correlations ranged from .41 to .93. 90% of the correlations were significant at the .01 level. The remaining 10% were significant at the .05 level. The general level of agreement for the Constructs item was high. Intra-class correlations for Item 2 are presented in Table 6.
Table 6

*Interrater Intra-class Correlations for Item 2: Constructs*

<table>
<thead>
<tr>
<th></th>
<th>Expert 1</th>
<th>Expert 2</th>
<th>Expert 3</th>
<th>Expert 4</th>
<th>Expert 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Expert 1</td>
<td>-</td>
<td>.696(**)</td>
<td>.722(**)</td>
<td>.928(**)</td>
<td>.550(**)</td>
</tr>
<tr>
<td>Expert 2</td>
<td>-</td>
<td>-</td>
<td>.723(**)</td>
<td>.800(**)</td>
<td>.633(**)</td>
</tr>
<tr>
<td>Expert 3</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>.723(**)</td>
<td>.408(*)</td>
</tr>
<tr>
<td>Expert 4</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>.633(**)</td>
</tr>
<tr>
<td>Expert 5</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

**Correlation is significant at the 0.01 level (2-tailed). *Correlation is significant at the 0.05 level (2-tailed).**

The Cronbach-Alpha coefficient for total scores across raters for the Use item was .91. Inter-rater correlations ranged from .56 to .75. All of the correlations were significant at the .01 level. The general level of agreement among raters for the Use item was high, and these figures are presented in Table 7.

Table 7

*Interrater Intra-class Correlations for Item 3: Use*

<table>
<thead>
<tr>
<th></th>
<th>Expert 1</th>
<th>Expert 2</th>
<th>Expert 3</th>
<th>Expert 4</th>
<th>Expert 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Expert 1</td>
<td>-</td>
<td>.696(**)</td>
<td>.748(**)</td>
<td>.711(**)</td>
<td>.655(**)</td>
</tr>
<tr>
<td>Expert 2</td>
<td>-</td>
<td>-</td>
<td>.651(**)</td>
<td>.807(**)</td>
<td>.631(**)</td>
</tr>
<tr>
<td>Expert 3</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>.558(**)</td>
<td>.618(**)</td>
</tr>
<tr>
<td>Expert 4</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>.662(**)</td>
</tr>
<tr>
<td>Expert 5</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

**Correlation is significant at the 0.01 level (2-tailed).**

The Cronbach-Alpha coefficient for total scores across raters for the Interpretation item
was .91. Inter-rater correlations ranged from .58 to .75. The intra-class correlations for this item are presented in Table 8. All of the correlations were significant at the .01 level. The general level of agreement among raters for the Interpretation item was high.

Table 8
*Interrater Intra-class Correlations for Item 4: Interpretation*

<table>
<thead>
<tr>
<th></th>
<th>Expert 1</th>
<th>Expert 2</th>
<th>Expert 3</th>
<th>Expert 4</th>
<th>Expert 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Expert 1</td>
<td>- .654(**)</td>
<td>.643(**)</td>
<td>.749(**)</td>
<td>.753(**)</td>
<td></td>
</tr>
<tr>
<td>Expert 2</td>
<td>- .578(**)</td>
<td>- .616(**)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Expert 3</td>
<td>- .663(**)</td>
<td>.581(**)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Expert 4</td>
<td>- .625(**)</td>
<td>- .625(**)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Expert 5</td>
<td>- .625(**)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Correlation is significant at the 0.01 level (2-tailed).**

The Cronbach-Alpha coefficient for total scores across raters for the ILP item was .91. Inter-rater correlations ranged from .48 to .75. All of the correlations were significant at the .01 level, and they are presented for this item in Table 9. The general level of agreement among raters for the ILP item was high.

Table 9
*Interrater Intra-class Correlations for Item 5: ILP*

<table>
<thead>
<tr>
<th></th>
<th>Expert 1</th>
<th>Expert 2</th>
<th>Expert 3</th>
<th>Expert 4</th>
<th>Expert 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Expert 1</td>
<td>- .476(**)</td>
<td>.752(**)</td>
<td>.672(**)</td>
<td>.700(**)</td>
<td></td>
</tr>
<tr>
<td>Expert 2</td>
<td>- .618(**)</td>
<td>- .571(**)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Expert 3</td>
<td>- .647(**)</td>
<td>.715(**)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Expert 4</td>
<td>- .830(**)</td>
<td>- .830(**)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Expert 5</td>
<td>- .830(**)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Correlation is significant at the 0.01 level (2-tailed).**
The Cronbach-Alpha coefficient for total scores across raters for the Intervention item was .88. Inter-rater correlations ranged from .43 to .72. 90% of the correlations were significant at the .01 level. The remaining 10% were significant at the .05 level. The general level of agreement for this item was high. Intra-class correlations for the Intervention item are presented in Table 10.

Table 10

<table>
<thead>
<tr>
<th></th>
<th>Expert 1</th>
<th>Expert 2</th>
<th>Expert 3</th>
<th>Expert 4</th>
<th>Expert 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Expert 1</td>
<td>-</td>
<td>.676(**)</td>
<td>.571(**)</td>
<td>.719(**)</td>
<td>.570(**)</td>
</tr>
<tr>
<td>Expert 2</td>
<td>-</td>
<td>.599(**)</td>
<td>.663(**)</td>
<td>.576(**)</td>
<td></td>
</tr>
<tr>
<td>Expert 3</td>
<td>-</td>
<td>.748(**)</td>
<td>.427(*)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Expert 4</td>
<td>-</td>
<td></td>
<td>.600(**)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Expert 5</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td>-</td>
</tr>
</tbody>
</table>

** Correlation is significant at the 0.01 level (2-tailed). * Correlation is significant at the 0.05 level (2-tailed).

Reliability Question 2

To what extent are the items that comprise the global AMAC-CTI measurement internally consistent?

An analysis of the influence of each item revealed that each is important. After removing each item to gauge the relative importance of each item, Cronbach-Alpha coefficients ranged from .93 to .94 and the results of this analysis are presented in Table 11. The overall Cronbach-Alpha for all items was .95.
Table 11: Assessment of AMAC-CTI Items

<table>
<thead>
<tr>
<th>Item</th>
<th>Correlation to Overall</th>
<th>Cronbach Post Item Removal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Purpose</td>
<td>.83</td>
<td>.93</td>
</tr>
<tr>
<td>Constructs</td>
<td>.75</td>
<td>.94</td>
</tr>
<tr>
<td>Use</td>
<td>.85</td>
<td>.93</td>
</tr>
<tr>
<td>Interpret</td>
<td>.89</td>
<td>.93</td>
</tr>
<tr>
<td>ILP</td>
<td>.85</td>
<td>.94</td>
</tr>
<tr>
<td>Intervention</td>
<td>.89</td>
<td>.93</td>
</tr>
</tbody>
</table>

Reliability Question 3
To what extent was the trait measured by the AMAC-CTI uni-dimensional?

An exploratory factor analysis of the AMAC-CTI scale was conducted to establish its uni-dimensional structure and provide an opportunity to inspect factor loadings. The specifications of this analysis were principal components with varimax rotation. This analysis produced a single factor with an Eigen value of 2.095 and accounted for 72.387% of variance. Only one component was extracted and the solution could not be rotated. This confirms that the AMAC-CTI is uni-dimensional. The results of this analysis are presented in Table 12.

Table 12: Eigen Values and Percent Variance

<table>
<thead>
<tr>
<th>Component</th>
<th>Factor Loading</th>
<th>Eigen Values</th>
<th>Percent Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>.140</td>
<td>2.095</td>
<td>72.4</td>
</tr>
<tr>
<td>2</td>
<td>.121</td>
<td>.300</td>
<td>10.4</td>
</tr>
<tr>
<td>3</td>
<td>.209</td>
<td>.186</td>
<td>6.4</td>
</tr>
<tr>
<td>4</td>
<td>.248</td>
<td>.140</td>
<td>4.8</td>
</tr>
<tr>
<td>5</td>
<td>.278</td>
<td>.104</td>
<td>3.6</td>
</tr>
</tbody>
</table>
Study 3 – Task Difficulty

To assess the level of difficulty associated with the performance task within this research, three inquiry questions were posed regarding task difficulty: 1) To what extent was the task challenging to test takers?; 2) To what extent was there variability in the performances of test takers?; and 3) To what extent was there evidence of lack of effortful processing or maximum performance?

Inquiry Question 1
To what extent was the task challenging to test takers?

Mean scores on the survey portion of the task ranged from 1.4 to 1.7. All of the mean scores were at the concrete or transitional level. The descriptive statistics for this inquiry question are presented in Table 13.

Table 13 Descriptive Statistics Regarding Task Performance

<table>
<thead>
<tr>
<th></th>
<th>Mean(^a)</th>
<th>Standard Deviation</th>
<th>Range</th>
<th>Skewness</th>
<th>Kurtosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall</td>
<td>1.5</td>
<td>.58</td>
<td>2.7 (.3 to 3.0)</td>
<td>.30</td>
<td>.05</td>
</tr>
<tr>
<td>Purpose</td>
<td>1.7</td>
<td>.60</td>
<td>2.6 (.4 TO 3.0)</td>
<td>.68</td>
<td>.38</td>
</tr>
<tr>
<td>Constructs</td>
<td>1.4</td>
<td>.59</td>
<td>3.0 (0 TO 3.0)</td>
<td>.27</td>
<td>.24</td>
</tr>
<tr>
<td>Use</td>
<td>1.5</td>
<td>.71</td>
<td>3.0 (0 TO 3.0)</td>
<td>.10</td>
<td>.44</td>
</tr>
<tr>
<td>Interpret</td>
<td>1.6</td>
<td>.76</td>
<td>2.8 (.2 TO 3.0)</td>
<td>.12</td>
<td>.70</td>
</tr>
<tr>
<td>ILP</td>
<td>1.4</td>
<td>.82</td>
<td>3.0 (0 TO 3.0)</td>
<td>.02</td>
<td>.77</td>
</tr>
<tr>
<td>Interventions</td>
<td>1.7</td>
<td>.65</td>
<td>2.8 (.2 TO 3.0)</td>
<td>.05</td>
<td>.57</td>
</tr>
</tbody>
</table>

\(^a\) (1=Concrete, 2=Transitional, 3=Formal)
Inquiry Question 2
To what extent was there variability in the performances of test takers?

The vast majority of participants were rated as concrete or transitional, with some variability among individual items. Overall, the raters rated 49% of the participants as concrete, 45% as transitional, and only 6% as formal. Participants appeared to have the greatest difficulty when attempting to define the constructs of the CTI. 82% of respondents were at the concrete level of this item. When you combine the findings in Table 15 below, 61% of respondents were operating at a transitional or formal operations level when asked to elaborate on the types of interventions they prescribe when using the CTI. The results of this analysis are presented in Table 14.

Table 14
Percent of Persons in Each Competency Level

<table>
<thead>
<tr>
<th></th>
<th>Concrete (0 To 1.4)</th>
<th>Transitional (1.5 To 2.4)</th>
<th>Formal (2.5 To 3.0)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall</td>
<td>49%</td>
<td>45%</td>
<td>6%</td>
</tr>
<tr>
<td>Purpose</td>
<td>42%</td>
<td>46%</td>
<td>12%</td>
</tr>
<tr>
<td>Constructs</td>
<td>82%</td>
<td>15%</td>
<td>3%</td>
</tr>
<tr>
<td>Use</td>
<td>58%</td>
<td>33%</td>
<td>9%</td>
</tr>
<tr>
<td>Interpret</td>
<td>49%</td>
<td>36%</td>
<td>15%</td>
</tr>
<tr>
<td>ILP</td>
<td>58%</td>
<td>30%</td>
<td>12%</td>
</tr>
<tr>
<td>Interventions</td>
<td>39%</td>
<td>49%</td>
<td>12%</td>
</tr>
</tbody>
</table>

Inquiry Question 3
To what extent was there evidence of lack of effortful processing or maximum performance?

From a possible total number of ratings of nine hundred and ninety (33 participants x 6
items x 5 raters), ninety-six ratings of 0 were put forth by raters (see Table 15). This means that 9.5% of the responses provided by survey participants were considered lacking any amount of rate worthy information. The distribution of these ratings ranged from five (purpose item) to thirty-two (ILP item). A review of the raw data revealed that all of the responses that received a zero rating were no more than three sentences long. The total number of ratings of 0 or 1 equaled 526. This suggests that 53% of the responses provided by doctoral level clinicians, doctoral graduate students, and master’s students who are trained in the use of the CTI and on average have administered ninety-two Career Thoughts Inventories were deemed as low to completely void of noteworthy information.

Table 15

<table>
<thead>
<tr>
<th>Rating of Zero</th>
<th>Total Ratings of Zero (n = 18)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Purpose</td>
<td>5</td>
</tr>
<tr>
<td>Constructs</td>
<td>19</td>
</tr>
<tr>
<td>Use</td>
<td>16</td>
</tr>
<tr>
<td>Interpret</td>
<td>16</td>
</tr>
<tr>
<td>ILP</td>
<td>32</td>
</tr>
<tr>
<td>Intervention</td>
<td>8</td>
</tr>
</tbody>
</table>

Study 4 – Convergent Validity

The intent of this study was to determine to what extent was task performance ratings as measured by the AMAC-CTI instrument correlated with supervisor ratings using the same theoretical model of measurement. The correlation between AMAC-CTI ratings and supervisor ratings was .31. This correlation was not significant. The mean AMAC-CTI rating was 1.7 with a standard deviation of .51, and the mean supervisor rating was 2.1 with a standard deviation of .58. The AMAC-CTI ratings are based on the average rating across all 6 items for all raters. Overall ratings ranged from 0 to 3. Supervisor ratings could be 1, 2, or 3. The results of this
A paired sample t-test was performed to examine whether there was a significant difference between AMAC-CTI ratings and supervisor ratings. The difference between these two samples was significant at the .05 level (t = 2.6; p = .02; df = 17, effect size = .734). In
conclusion, based on the findings of the correlation analyses and this t-test, concurrent validity between AMAC-CTI ratings and supervisor ratings was not demonstrated.

**Study 5 – Convergent Validity: Correlates of Task Performance Ratings as Measured by the AMAC-CTI**

The fifth study attempts to answer the research question, to what extent were ratings on the AMAC-CTI associated with education level, the number of assessment courses completed, the number of Career Thoughts Inventories administered, and the number of Career Thoughts Inventories review with a supervisor. The results of this analysis are presented in Table 17. Pearson product correlations revealed that ratings from the AMAC-CTI were significantly correlated with the number of CTIs that participants had reviewed with their supervisors. The number of CTIs administered and understanding of the purpose and constructs of the CTI were also significantly related. The amount of coursework in assessment classes and research design appeared inversely related to the AMAC-CTI instrument but the measure of these correlates was not significant. The amount of coursework in assessment and research design was not related to scores earned on the AMAC-CTI.

**Table 17**

*Correlation: Education, Experience, and AMAC-CTI Scores (n = 33)*

<table>
<thead>
<tr>
<th></th>
<th>Education</th>
<th>Coursework</th>
<th>CTI Admin</th>
<th>CTI Review</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall</td>
<td>.31</td>
<td>-.10</td>
<td>.34</td>
<td>.47**</td>
</tr>
<tr>
<td>Purpose</td>
<td>.13</td>
<td>-.05</td>
<td>.40*</td>
<td>.46**</td>
</tr>
<tr>
<td>Constructs</td>
<td>.28</td>
<td>-.04</td>
<td>.50**</td>
<td>.49**</td>
</tr>
<tr>
<td>Use</td>
<td>.33</td>
<td>-.29</td>
<td>.22</td>
<td>.27</td>
</tr>
<tr>
<td>Interpret</td>
<td>.24</td>
<td>-.08</td>
<td>.16</td>
<td>.35*</td>
</tr>
<tr>
<td>ILP</td>
<td>.35*</td>
<td>-.06</td>
<td>.20</td>
<td>.39*</td>
</tr>
<tr>
<td>Interventions</td>
<td>.18</td>
<td>.06</td>
<td>.30</td>
<td>.47**</td>
</tr>
</tbody>
</table>

** significant at .01 level; *significant at .05 level.
While the findings the relationship between CTI scores and Experience (CTI Administrations and number of CTIs administered and reviewed with a supervisor) was encouraging, there was a noticeable gap between one individual and the remaining thirty two participants. A correlation analysis using a square root derivation was performed to examine the magnitude of this bias (see Table 18). After this adjustment, a mild shift in magnitude and significance levels across the correlation matrix was observed.

Table 18

_Correlation and Descriptive Statistics: Square Root Experience (n = 33)_

<table>
<thead>
<tr>
<th></th>
<th>SQ Admin</th>
<th>SQ Review</th>
<th>Skewness</th>
<th>Kurtosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall</td>
<td>.32</td>
<td>.42*</td>
<td>.28</td>
<td>.13</td>
</tr>
<tr>
<td>Purpose</td>
<td>.40*</td>
<td>.43*</td>
<td>.68</td>
<td>.38</td>
</tr>
<tr>
<td>Constructs</td>
<td>.51**</td>
<td>.43*</td>
<td>-.27</td>
<td>2.4</td>
</tr>
<tr>
<td>Use</td>
<td>.15</td>
<td>.17</td>
<td>.10</td>
<td>-.44</td>
</tr>
<tr>
<td>Interpret</td>
<td>.16</td>
<td>.36*</td>
<td>-.12</td>
<td>-.70</td>
</tr>
<tr>
<td>ILP</td>
<td>.14</td>
<td>.32</td>
<td>.02</td>
<td>-.77</td>
</tr>
<tr>
<td>Interventions</td>
<td>.32</td>
<td>.45**</td>
<td>-.05</td>
<td>-.57</td>
</tr>
</tbody>
</table>

** significant at .01 level; *significant at .05 level.
CHAPTER 5
DISCUSSION

Introduction

This study was designed to validate a measure of counselor competency in the use of psychological assessments. The name of the measure created is A Measure of Assessor Competency for the Career Thoughts Inventory (AMAC-CTI). The AMAC-CTI is based on a theoretical construct called the Piagetian Matrix of Test User Competence (PMTUC). The PMTUC represents a merger of Jean Piaget’s theory of cognitive development and a constellation of studies regarding competency development (Stoltenberg, 2005; Stoltenberg et al, 1994; Blustein et al, 1994; Jennings & Skovholt, 1999; Krishnamurthy et al, 2004). The AMAC-CTI is designed to differentiate among three levels of competency in the use of psychological assessment. These three levels are concrete, transitional, and formal operations. Since the AMAC-CTI is a measure of competency in the use of assessments, securing experts of a given psychological assessment was important. The Career Thoughts Inventory (CTI) was selected for this study because the creators of the CTI and clinicians that have extensive experience in the use of the CTI were accessible for this research.

Students and professionals in the field of counseling psychology with experience in the use of the CTI were asked to volunteer for this study. The study required participants to complete a web-based survey divided into two parts. The first part of the survey collected data on the education of the respondent, the number of assessment theory courses each respondent had taken, the number of CTIs administered, and the number CTIs each subject had administered and reviewed with a supervisor. The second part of the web-based survey included 6 questions. Each question represented an item or construct of interest.

Findings and Conclusions

As noted above, the intent of this research was to validate the AMAC-CTI. Five studies were conducted to achieve this end. The first study, content validation, revealed that five experts in the field of counseling and career development unanimously agreed that all six items that comprise the AMAC-CTI were valid measures of competency and either important or critically
important components of competency measurement. All the experts agreed that knowledge of the purpose and constructs of a psychological assessment was critically important. Therefore, the conclusion is that the six items of the AMAC-CTI are important components in the measurement of assessor competence in the use of the Career Thoughts Inventory.

The second study, reliability, consisted of three inquiry questions. The first inquiry regarding inter-rater reliability demonstrated rater consistency when using rating participant performance. The Cronbach-Alpha of the overall score for raters was .95. The scoring rubrics based on Piagetian levels of development worked very well as evidenced by the strong inter-rater reliability coefficients across all items and between the items and overall AMAC-CTI rating of competency. Regarding the second inquiry question, an analysis of the influence of each item revealed that each item contributes to the internal consistency of the AMAC-CTI. The third inquiry question regarding the uni-dimensionality of the AMAC-CTI was addressed using an exploratory factor analysis (EFA). This analysis confirmed that the AMAC-CTI is uni-dimensional. The analysis produced a single factor with an Eigen value of 2.095 and accounted for approximately 73% of the variance. Therefore, the conclusion is that the AMAC-CTI not only demonstrated internal consistency but it produced a uni-dimensional measurement.

The third study, item difficulty, was developed to gauge the difficulty of the survey performance task. Three inquiry questions were addressed in this study. Inquiry question one examined mean scores and distribution patterns to determine the extent in which this task was difficult. Based on mean ratings, the results suggest that the performance task was somewhat difficult. Mean scores ranged from 1.4 to 1.7. These scores are consistent with the findings from a field study conducted in 2004 (Etheridge, Peterson, & Hill). Skewness and kurtosis statistics indicate that the distribution of performance for each item and the overall score were normally distributed. Inquiry question two examined the extent to which there was variability in the performances of test takers. In Piagetian terms, the majority of participants (94%) were rated as concrete (face value interpretations completely tied to the instrument and constrained by reality or context) or transitional (able to identify a problem and use the appropriate assessment). Only six percent of the participants demonstrated knowledge considered to be at a formal level using the PMTUC model of measurement, meaning that the participants integrated the instrument using feedback, dialogue, the results of the instrument, and the understandings that arise of each of these three perspectives. Participants appeared to have the greatest difficulty when attempting
to define the constructs of the CTI (82% at the concrete level). Respondents appeared quite competent when asked to elaborate on the types of interventions they prescribe when using the CTI (39% at the concrete level). Inquiry question three examined the extent to which there was evidence of effortful processing or exerting a maximum performance on a task. From a possible total number of ratings of 990 ratings, there were 96 ratings of zero. A score of zero may have indicated that a given rater felt the response a given survey participant offered completely failed to answer the AMAC-CTI item. This means that 9.5% of the responses provided by survey participants were below the quality needed to receive a rating of one, the lowest rating possible. A review of the raw data revealed that all of the responses that received a zero rating were no more than three sentences long. Since the majority of participants were doctoral level clinicians or third year doctoral students well trained in the use of the CTI, these findings suggest there was either a problem with effortful processing or maximum performance. Another possible reason is the participant lacked sufficient knowledge to respond but given the average level of education of participants coupled with the level of experience participants possessed in the use of the CTI (average administrations = 92), lack of sufficient knowledge seems unlikely. Therefore, the conclusions for study 3 were that the task was challenging, that nearly all participants fell into the concrete or transitional level of competency, and that maximum performance was not obtained throughout the test by all participants.

The fourth study, convergent validity, examined the extent to which task performance ratings as measured by the AMAC-CTI instrument correlated with supervisor ratings using the same evaluation criteria. The correlation between AMAC-CTI ratings and supervisor ratings was $r = .31$ with a $p > .05$. The mean AMAC-CTI rating was 1.7 whereas the mean supervisor rating was 2.1. A t-test revealed that the difference between the mean scores was significant at the .05 level ($p = .02$). Therefore, there was no evidence to support the convergent validity of the AMAC-CTI in terms of a relationship with supervisor ratings.

The fifth study examined the extent in which ratings on the AMAC-CTI were associated with 1) education level, 2) the number of assessment courses completed by survey respondents, 3) the number of Career Thoughts Inventories administered, and 4) the number of Career Thoughts Inventories administered and reviewed with a supervisor. Analyses revealed that overall ratings from the AMAC-CTI were significantly correlated with the number of CTIs that participants reported reviewing with their supervisors ($r = .48$, $p < .01$). The number of CTIs
reviewed with supervisors was also correlated with the purpose item ($r = .46, p<.01$), the constructs item ($r = .49, p<.01$), interpretation item ($r = .35, p < .05$), ILP item ($r = .39, p < .05$), and the intervention item ($r = .47, p<.01$). The number of CTIs administered was correlated with the item understanding the purpose of the CTI ($r = .40, p<.05$) and the item understanding the constructs of the CTI ($r = .50, p<.01$). Education level and the number of courses in assessment were not consistently correlated with competency in the use of the CTI as measured by the AMAC-CTI. The findings were consistent with the idea that experience with the use of a given psychological assessment is important to competency.

Limitations of the Study

The external validity or generalizability of the results of this research is influenced by several limitations. First, this attempt to validate a measure of competency is based on the assumption that the six items selected for the determination of competency are correctly identified. Despite the fact that the creators of the AMAC-CTI borrowed from existing literature when building the AMAC-CTI, that the AMAC-CTI was field tested, and that experts in the field of career counseling rated the items within the AMAC-CTI as important to critical in measuring competency, there is still a possibility that the creators failed to identify a critical item in the measurement of competency. The second limitation is that the AMAC-CTI was created to measure assessor competence in the use of just one psychological assessment. At the conclusion of this dissertation, the results might be generalizable only to persons that use the Career Thoughts Inventory in career counseling settings. The experts and participants in this study were student counselors or clinicians that possess a significant training history in career counseling. The third limitation is tied to the second limitation. All of the participants and experts were tied to the same training program. All participants were enrolled in graduate studies in the Counseling Psychology Program at a large southeastern university, were graduates of the counseling program at FSU, were faculty members within the Combined Doctoral Program in Counseling Psychology and School Psychology at a large southeastern university at the time, or were professional staff at the FSU Career Counseling Center. The fourth limitation of this research is the lack of diversity among participants. While the demographic distribution of participants was consistent with the employment and enrollment demographic of the graduate program, greater representation of minorities would be desirable. The fifth limitation to the
generalizability of this research is the fact that all the participants were volunteer participates and willing to complete a survey response task that took between 30 and 75 minutes to complete.

Sixty-eight (68) eligible candidates were contacted about participating and 39 agreed. Six of the thirty-nine participants started the survey but did not complete it. The sixth limitation is sample size. The sample of 33 survey respondents and only 18 student respondents for the convergent validity study are too few participants to assert strong conclusions.

While the limitations above undoubtedly impact generalizability, there are two principal limitations of this research that may threaten the internal validity of this dissertation. First and foremost, do the responses from the performance task (answering the six questions about instrument use) represent the actual amount of knowledge participants possess? Study 3 was designed to assess the limitation regarding the likelihood that the responses to the six items that comprise the AMAC-CTI represented maximum performance. Frequency analyses indicated that almost ten percent of the sample provided responses that were rated as zero. A review of the responses showed that these zero rated responses were typically two to three sentence long.

While not impossible, it is difficult to describe the selection process clinicians employ when choosing an assessment, defining the constructs of the assessment instrument, or the steps one takes when creating a treatment plan using results from an assessment in two sentences. To add, 53% of the responses from well-trained and highly educated clinicians who report administering on average, ninety-two CTIs, fell into the concrete or lack of effortful processing (rating of 0) category.

The problem of maximum performance or lack of maximum performance did not impact Study 1 regarding content validity. Moreover, sub-maximum performance may only slightly impact the analysis of inter-rater reliability in Study 2. Studies 3, 4, and 5 are significantly impacted. The determination of task difficulty is highly biased by poor performance. Sub-maximum ratings load on the conclusion that the performance task is difficult. Study 4, convergent validity, compares AMAC-CTI ratings to supervisor ratings of each student respondent’s demonstrated competency on the job. If student participants have a history of competence in the use of the CTI on the job but do not take this research’s performance task seriously, the on-the-job supervisor ratings for these students will differ greatly to the ratings on the performance task. A comparison of supervisor ratings and AMAC-CTI ratings revealed that this occurred on several occasions. Several students that received high ratings from supervisors
offered “thin” or two to four sentence responses on some items that resulted in reduced scores. The differences in the means also reflect the fact that supervisor ratings of students tended to be higher than student performance as measured by the AMAC-CTI. Mean AMAC-CTI scores were 1.7 compared to mean supervisor ratings of 2.1.

Study 5 examined the relationship between education and experience and AMAC-CTI ratings. These findings are also susceptible to sub-maximum performance. This validation research expected to find correlations between education and AMAC-CTI scores and experience and CTI scores. No significant relationship between education level and AMAC-CTI scores was noted. A review of the raw data revealed that a sizeable percentage of doctoral level clinicians (highest education level possible) offered thin or irrelevant responses.

This limitation might be addressed by adding weight to the importance of the response task. This might range from having students complete the survey for a grade to having students complete this task with the knowledge that their supervisors would be reading their responses. Another way to address this limitation might be to move from a web-based approach to a face-to-face interview. Responses from the earlier pilot study (Etheridge et al, 2004) were collected via face-to-face interview and this survey format yielded higher levels of responses. Finally, consideration of task difficulty is warranted. By examining the PMTUC, AMAC, or recent literature that addresses task performance problems, perhaps the response format could be altered to reduce task difficulty.

The second principal limitation of this research is the number of supervisors that participated in the convergent validity study. The five expert raters that participated in this study were all eligible supervisors. However, of all eighteen students that volunteered to participate in this research, sixteen were supervised by one of the experts. The remaining two volunteers were supervised by a master’s level clinician that was not an expert. This means that a preponderance of supervisor ratings was provided by one supervisor. This does not provide the variation necessary for this study of convergent validity to be valid. Another limitation associated with the number of supervisors is the error associated with an overly generous supervisor rater. With additional raters, the problem of inflated ratings may persist but there are statistical procedures to address this possible problem when a sufficient number of raters are available. With only two raters (one primary rater), the odds that any attempted measure of the relationship between AMAC ratings and supervisor ratings would be overly influenced by bias are increased.
As in all research, some limitations are more critical than others. Related to this study, the most critical limitations are maximum performance followed by the limitation of sample size, lack of an adequate number of supervisor ratings, lack of minority participants, restricted geographic range of experts and participants, restricted subdiscipline range, restricted instrument, and the fact that all participants were voluntary participants. As noted above, these limitations can be categorized into two distinct tiers best described as a critical limitation tier and a limitation of concern. The problem with participant maximum performance, adequate sample size, and adequate number of supervisor ratings for the convergent validity study are critical limitations of this study.

Implications for Practice

This research explored the feasibility of validating a competency-rating instrument that measures training progress in the use of assessment instruments in career counselor training. The broader intent of this endeavor was to create a measure of assessor competence in the use of any psychological assessment instrument in the field of psychology. The AMAC could be viewed as a flexible measure that can be adapted and applied to any psychological assessment or training paradigm. Differentiating between master therapists and non-masters is not the intent of this research. This dissertation is intended to further examine the possibility of validating a measure of assessor competency by validating an instrument that attempts to rate abstract thought processes.

One implication for practice could be the utility of the AMAC in graduate preliminary exams. Graduate programs could use the AMAC template to build a standardized measurement of responses to preliminary exam questions. The AMAC can be used to assess levels of competency in the use of psychological assessments or it can be adjusted to examine the extent in which student counselors are becoming competent in the adoption of a given graduate training program’s clinical treatment philosophy. Graduate programs already use preliminary exams to ascertain the general level of knowledge of students. Moreover, questions on preliminary exams often assess the items that collectively comprise the AMAC total score of competency. Students are asked to elaborate on their knowledge about the theories that direct their clinical actions (purpose and construct items), the steps they take when using these theories (use item), how their interpretations are influenced by theories (interpretation item), and how they develop treatment
plans and intervention strategies based on these theories (ILP and intervention items). The AMAC is a possible blueprint for the standardization of measuring abstract thought processing as measured through preliminary exams.

However, as noted above, the AMAC was created with the intent to remediate competency problems. Therefore, the intent of the creators of the AMAC is not for training programs to wait until preliminary exams to employ the AMAC. The AMAC is ideally suited for every practicum and/or every psychological assessment a student clinician might encounter. The overall rating of competency is designed to inform counselor trainers as to the overall operating level of student counselors but the items on the AMAC are designed to highlight the specific area of weakness. For example, a student counselor may be rated by assessment course instructor or practicum supervisor as operating at a concrete level in the use of the WISC-IV. Upon review of item ratings, the student may illustrate a strong grasp of the content items on the WISC-IV and how to interpret the findings of the instrument but have difficulty in constructing an individualized learning plan for clients. As noted in the chapter one of this dissertation, the majority of the governing bodies of practicing psychologists have promulgated rules and standards pertaining to the appropriate use of psychological assessments (American Association for Marriage and Family Therapy, 1985; American Counseling Association, 1995; American Educational Research Association, 1985; American Psychological Association, 1999; National Association of School Psychologists, 1992; National Council on Measurement in Education, 1995). While the original emphasis of test practices focused on psychometric knowledge, recent standards emphasize the need for test users to be skilled and knowledgeable about specific contexts and rationales of use. Another statement put forth in chapter one and worth repeating is, regardless of the intent of any governing body, an outline of standards is only as effective as the tools available to measure adherence to those standards (Etheridge, Peterson, & Hill 2004). The development of instruments like the AMAC would allow graduate training programs to gauge the level of competency counselors demonstrate when using the assessments they are being trained to use.

The specific implications for practice of the findings of this research (AMAC-CTI) is that student counselors who use the CTI when providing counseling services to clients at the a large southeastern university Career Counseling Center may need further training in the use of the CTI. Also, the faculty that provides CTI training to student counselors at the large southeastern
Implications for Further Research

Future research regarding the AMAC-CTI could focus on building measures of assessor competency across the broader areas of clinical, counseling, and school psychology. The AMAC could be viewed as a flexible measure that can be applied to any psychological assessment or training paradigm. Given this, research on the feasibility of creating a measure of assessor competency in the use of the MMPI-2, the Rorschach, the Wechsler scales, and other clinical tools could be explored. Equally stated, research on the feasibility of building PMTUC models and related AMAC-CTI tools in the measurement of the relative application of counselor training program ideologies should be explored. Also, the AMAC-CTI questions could be applied to a stimulus employing case material. The present stimulus of the AMAC-CTI called for hypothetical abstract reasoning about the nature of the CTI, thus making the task more difficult. The incorporation of case material may bring down the difficulty level.

Future research must address the problem with comparing AMAC ratings with supervisor ratings. The AMAC ratings were strictly criterion based while supervisor ratings were normative. Future methodology should have more raters with a more refined instrument rating measure for practicum performance. The AMAC is based on the Piagetian Matrix of Test User Competency. Future research may need to add to the literature base in this text to examine the efficacy of the PMTUC as a guiding theory. Perhaps the problems with the complexity of the task and convergent validity are due to the inadequacy of the PMTUC.

It is essential that future research attempt to expand the sample sizes to better reflect the diversity in the profession and to improve the power of the findings. Additionally, sample size undoubtedly influenced the exploratory factor analysis performed. When performing factor analyses, minimum sample sizes are recommended between 150 and 300 subjects (Guadagnoli & Velicer, 1988; Tabachnick & Fidell, 1996). Future research that seeks to build on these specific findings should attempt to reduce the level of difficulty of the performance task, find raters that are well-skilled in the use and supervision of the use of the CTI that are not directly tied to the
instrument, and conduct convergent validity analyses with at least three sets of supervisor ratings to compare with AMAC-CTI scores. On-site research could focus on using the results of this research to provide consultation services regarding the revision of the current CTI training curriculum. Upon revision and application of the new curriculum, new counselor trainees could be assessed using the AMAC-CTI. Cohort scores from this AMAC-CTI application could be compared to this research and the field test results conducted in 2004.

Summary and Discussion of Results

This research provides evidence toward the validation of the AMAC-CTI. Content validity was established based on expert ratings of the content. Feedback regarding the importance of the content items of the AMAC-CTI should be sought from experts that are not directly or indirectly involved in the creation of the assessment instrument selected, the CTI. Inter-rater consistency and internal consistency was demonstrated in study 2. The EFA performed in this study demonstrated that the AMAC-CTI is uni-dimensional. Study 3 revealed that the performance task is somewhat difficult. Future research could attempt to reduce this difficulty. Study 4, the convergent validity study that examined the relationship between AMAC-CTI ratings and supervisor ratings failed to demonstrate convergent validity. Two problems with this attempt to demonstrate convergent validity were that one supervisor provided 89% of the supervisor ratings and only 18 pairs of ratings were available. Even if a relationship between supervisor ratings and AMAC-CTI ratings was unearthed, with only 18 participants in this study, there was insufficient power to demonstrate convergent validity. Study 5 revealed that a significant positive relationship was present between AMAC-CTI total scores and the number of CTIs participants reviewed with a supervisor. This is consistent with the idea that clinical supervision is critical to competency development.
APPENDIX A

PIAGETIAN MATRIX OF TEST USER COMPETENCE (THEORETICAL MODEL)
Table A1.

*Piagetian Matrix of Test User Competence (Theoretical Model)*

<table>
<thead>
<tr>
<th>Construct</th>
<th>Concrete</th>
<th>Transitional</th>
<th>Formal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Understanding of instrument</td>
<td>Limited: states dimensions only; empty or glib elaboration</td>
<td>States dimensions and underlying theory to practice link.</td>
<td>Explanation linked to diagnostic and prescriptive qualities of instrument.</td>
</tr>
<tr>
<td>Constructs of instrument</td>
<td>Names constructs only</td>
<td>Identifies constructs and connotes theoretical link.</td>
<td>Operationalizes the multidimensional links implied by the theory base.</td>
</tr>
<tr>
<td>Assignment of instrument</td>
<td>Contextually bound, vague rules; non-discriminatory</td>
<td>Appropriately identifies the rationale for use</td>
<td>Appropriate identification coupled with well-founded contextual integration.</td>
</tr>
<tr>
<td>Examples of Assignment of</td>
<td>Offers event or instance only; incongruent use</td>
<td>Event with parroted rationale; mildly congruent with theory.</td>
<td>Clear rationale with multiple events linked to abstract discrimination.</td>
</tr>
<tr>
<td>Reasons not to use instrument</td>
<td>Vague rules or non-relevant elaboration on specific event.</td>
<td>Clear but sweeping rules; hypergeneralized rationales.</td>
<td>Clear rules substantiated by contextual indications; case-by-case discrimination.</td>
</tr>
<tr>
<td>Construct</td>
<td>Concrete</td>
<td>Transitional</td>
<td>Formal</td>
</tr>
<tr>
<td>----------------------------</td>
<td>----------------------------------------------------</td>
<td>-----------------------------------------------------</td>
<td>-----------------------------------------</td>
</tr>
<tr>
<td>Example of non-assignment</td>
<td>Offers event or instance only; incongruent use</td>
<td>Specific events coupled with generalized procedural rules.</td>
<td>Free of generalized rules; less event specific; case-by-case abstraction.</td>
</tr>
<tr>
<td>Interpretive steps</td>
<td>Procedural only</td>
<td>Procedural fueled by results not just instrument.</td>
<td>Less procedural and more process oriented fueled by multiple sources.</td>
</tr>
<tr>
<td>Information from instrument</td>
<td>Procedural; conceptual relevance to client and future intervention missing, vague.</td>
<td>Less procedural; client oriented but lacks robust link of instrument theory and clinical practice.</td>
<td>Client centered with minimal procedure.</td>
</tr>
<tr>
<td>Intervention developed</td>
<td>Vague generalities; overly tied to instrument</td>
<td>Mildly decontextualized; results are on a continuum.</td>
<td>Uses client cues, instrument, and other available resources; dynamic.</td>
</tr>
</tbody>
</table>
APPENDIX B
VITA OF PRINCIPAL RESEARCHER
Vita

Roy L. Etheridge

Education:
Ph.D Candidate, Counseling Psychology, Florida State University, Tallahassee, FL. 1998 to present
M.S. in Psychology, 1998, Florida Atlantic University, Boca Raton, Florida (ABT).
B.S. in Psychology, 1996, Old Dominion University, Norfolk, Virginia.
Emergency Medical Technician, National Cert., 1987, School of Aerospace Medicine, Brooks AFB, TX

Recent Grants:
2006: Co-author BRITE grant (expand treatment capacity). FL received $14,000,000 award.
2004: Co-author Hurricane Recovery grant. FL received $11,000,000 award.
2004: Contributing author for Access to Recovery client voucher grant. Received $21,000,000 award.

Specialization/Interest Areas:
Statistical Applications and Empirical Design
Intervention Strategies for Clients in Chronic Vocational Distress
Personality Theory and Related Consultative Applications

Statistical Programs/Skill Level:
SPSS – Proficient
SYSTAT - Proficient
Minitab - Proficient
LISREL – Familiar
SAS – Familiar
MS Excel – Expert
MS Access - Proficient

Statistical Coursework:
Linear Regression
Multivariate Statistics
Hierarchical Linear Modeling
Structural Equation Modeling
Advanced ANOVA Models
Non-Parametric Statistics

Professional Experience:

Consultation and Research:

Coordinator, Statistical Research Florida Mental Health Institute, Assigned to Florida
Research Associate, Department of Children and Families Substance Abuse Program, Tallahassee, FL, August 2001 to November 2002. Primary duty was data analysis and data management for the 12-month follow-up survey. Other duties included data analyses and report writing for ad hoc assignments and development of a HIPAA library (Health Insurance Portability and Accountability Act of 1996).

Consultation and Statistical Services Specialist, Center for the Study of Technology in Career Counseling and Development, Tallahassee, FL, April 2000 to November 2002. Consultation services included assisting clients with data analyses, research design, and methodology. Statistical Services included data entry, data analyses, graphic design, and programming.

Research Assistant, Center for the Study of Technology in Career Counseling and Development, Florida State University, Tallahassee, FL, August 1998 to June 2001. Primary duties included data analyses, data management, and assisting with the development of research designs in career counseling.

Research Assistant, Florida Atlantic University, Boca Raton, FL, August 1996 to August 1998. Primary duties included data analyses, collection, and entry. Other duties included literature reviews, writing, and editorial reviews.

Research Assistant, Old Dominion University, Norfolk, VA. September 1995 to June 1996. Duties included data analyses, independent design, development, and pursuit of research investigating the effects of mass media portrayals of women on violence against women. Assisted with literature review, grant proposal, and research design. Assistantship funded by grant from NAACP.

Teaching and Training:

Group Processes Instructor, Human Services Center, Florida State University, Tallahassee, FL, June 2000 to September 2000. Prepared lectures on group counseling techniques and the variations associated with strategies across different populations.

Individual Development and Practicum Supervisor, Human Services Center, Florida State University, Tallahassee, FL, January 2000 to June 2000. Supervised Master's level counselor trainees enrolled in Practicum. Observed client sessions to provide feedback and guidance on basic techniques and mechanics. Assisted trainees with case
conceptualizations, clinical notations, diagnostic indicators, and treatment plans.

Assessed bi-monthly critiques of research publications in human services to monitor the progress of the analytic and critical evaluation skills of counseling graduate students. Organized, drafted, and administered final exam. Performed data analyses on final exam items.

**Counselor Trainer and Trainee Supervisor**, Human Services Center, Florida State University, Tallahassee, FL, August 1999 to January 2000.
Introduced first year graduate level trainees to professional counseling, the practicum environment, administrative and judicial mandates, and the ethical standards of practice. Created a pre-practicum training model that attends to the philosophical, personal, and professional convictions of each trainee in concert with the ethical, theoretical, and practical application principles of counseling theory through disclosure exercises, lecture, and experience.

Designed lectures on data analysis, the construction of applied statistical programming, and language modifications of the statistical software program Minitab.

Duties included assisting with daily instructional design and lecture on data analysis, applied statistical processes, and research design. Assignment, assessment, and overall maintenance of student evaluations. Duties required acquisition of fluent Minitab programming ability.

**Teaching Assistantship**. Developmental Psychology (PSY4164) course at Florida Atlantic University, Boca Raton, FL. May 1997 to August 1997.
Duties included weekly selection, design, and lecture on current topic relevant research publications and literature. Assist with exam administration, assessment, and maintenance of student performance records.

**Clinical Assessment and Counseling:**

**Senior Counselor**, Department of Corrections, Autry State Prison, Pelham, GA. October 2002 to March 2003.
Provided individual and group counseling services to inmates at medium security prison. Responsible for the monitoring, treatment, diagnostic management, bi-monthly review, and bi-annual evaluation of 60 level 3 mental health inmates. Other duties included providing GA DOC with clinical updates and evaluations on probation eligible inmates and providing instructional courses on mental health to correctional officers.

Review clinical and criminal files of sexual offenders referred from the Florida Department of
Corrections. Prepare written clinical summaries of individual’s clinical and criminal history and present each case to multidisciplinary team for determination of eligibility for civil commitment as a sexually violent predator. Collaborate with state attorney’s offices, the Dept. of Corrections, and other relevant agencies in gathering necessary information for case files.

**Psychological Assessment Specialist and Adjustment Counselor,** HealthSouth Rehabilitation Hospital, Tallahassee, FL. January 2001 to August 2001. Duties include assessing the cognitive and psychosociological status of patients that are admitted to this hospital for rehabilitation services and providing counseling services to inpatients that are suffering from a variety of psychological and sociological adjustment issues. Assess physiological, pharmacological, and neurobiological factors associated with psychosocial functioning of patients.

**Career Services Counselor,** Florida State University Career Center, Tallahassee, FL. August 2000 to June 2001. Provide career and mental health counseling services to clients in career decision making distress. Provide brief staff assisted services to clients through the management and application of the Center’s network of information systems, data files, and Computer Assisted Career Guidance resources.


**Psycho-Educational Assessment Specialist,** Center of Multidisciplinary Studies and Evaluation, Florida State University, Tallahassee, FL. May 2000 to August 2000. Administered, scored, and evaluated psychological, sociological, and psychoeducational assessment batteries for the purpose of intelligence testing, learning disability reviews, and educational placement services.

**Social Services Administrative Liaison,** Mercy American River Hospital, Sacramento, CA. September 1992 to December 1993. Primary duties included evaluating state and hospital intervention programs through medical record reviews. Duties involved coding program effectiveness ratings by summarizing treatment histories, clinical histories, and all other chronic diagnostic or symptomatic criteria.

**Contingency Counselor** during Gulf War. January 1991 to March 1991. Shared counseling duties with Army Chaplains and Military Debriefment Unit while on active duty as a paramedical specialist serving in the Gulf War. Treated Gulf War related stress and personal management disorders during the conflict.

**Special Assignment,** Contracted selection at McClellan AFB, Sacramento, CA. April 1990 to November 1990. Selected to assist contracted Skills Management Team in the administration of the MMPI to general office staffs on base. Duties included counseling (post test) office staff members on
awareness strategies developed using the MMPI to create perceptual relations models.

**Aerospace Medical Services Specialist,** USAF Clinic McClellan, Sacramento, CA. March 1987 to March 1991. Position held in active duty service for the United States Air Force. Responsibilities were 50% clinical, 25% administrative, and 25% instructional. Instructional duties including CPR training, wartime mobility training, and hazardous material training. Clinical duties included administering all necessary physical examination measures within scope of technical training.

**Other Experience:**

**Bartender,** TGI Fridays, Sacramento, CA; Norfolk, VA; Boca Raton, FL; Fort Lauderdale, FL; Tallahassee, FL. May 1991 to July 1999. Duties included customer service and satisfaction and maintenance of surrounding area. Worked for Fridays periodically while pursuing my undergraduate and graduate studies.

**Staff Writer,** The Mace and Crown, Old Dominion University, Norfolk, VA. August 1995 to July 1996. Duties including editorial reviews of articles submitted to the Mace and Crown and the weekly submission of a feature article that touches upon societal norms and current media topics of interest.

**AM Talk Radio Co-host,** WODU – AM 790, Norfolk, VA. January 1994 to December 1995. Discussed and debated merits of recent political developments and social reformations with my conservative co-host. Bi-weekly show broadcast to approximately 6,500 listeners.

**Manager,** Awful Arthurs Oyster Bar, Urbanna, VA. March 1994 to July 1994. Managerial duties included supervision of staff, personnel decisions, and scheduling. Responsible for tracking inventory and all related logistics.

**Political Negotiations Representative,** California League of Conservation Voters, Sacramento, CA. March 1991 to August 1992. Primary responsibilities included the update, review, and maintenance of legislative precedents, current referendums and legislation, and legislator profile composites. Other duties involved assisting in amendment drafts and all other lobby efforts.

**Research Projects:**

1999 to present: Completed a pilot study in the spring of 1999 that explored the use of the Career Thoughts Inventory and counselor competency in the training setting. Reproduced study with larger population and are currently reviewing findings. Assisted in the development and application of an evaluation model that is designed to assess the factors associated with the increase of competency as it relates to the use of assessments in career counseling. Currently validating Competency Model for dissertation.

Etheridge, R. (1996) *The Effects of Women Portrayals as Sex Objects on Male Sexual Attitudes.* Independent pursuit of research project supervised by Dr. Patricia Clark. Research development, design, and implementation independently initiated but further data is needed for conclusive results.


**Professional Presentations:**


Etheridge, R. (1994) *Attitudes and Condom Use.* Poster presented during annual presentation series sponsored by the Department of Psychology, Old Dominion University. Norfolk, VA.

**Papers to be submitted for publication:**

APPENDIX C
A MEASURE OF ASSESSOR COMPETENCE – CAREER THOUGHTS INVENTORY
RESPONDENT RATING FORM (THE SCORING MATRIX)
### AMAC-CTI RESPONDENT RATING FORM

**RATER NAME:** __________________________________________

**DATE:** ___ /___/2007

<table>
<thead>
<tr>
<th>Construct</th>
<th>Concrete Rating of 1</th>
<th>Transitional Rating of 2</th>
<th>Formal Rating of 3</th>
<th>Trainee Rating (Circle)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level in which the respondent describes the purpose of the CTI</td>
<td>Limited: aware of dimensions only; empty or glib elaboration; One sentence only.</td>
<td>Shows some understanding of dimensions and underlying theory to practice link.</td>
<td>Understanding fully linked to diagnostic and prescriptive qualities of instrument.</td>
<td>1</td>
</tr>
<tr>
<td>Level in which the respondent defines the constructs of the CTI</td>
<td>Able to identify constructs only; One sentence only.</td>
<td>Identifies constructs and connotes theoretical link.</td>
<td>Operationalizes the multidimensional links implied by the theory base.</td>
<td>3</td>
</tr>
<tr>
<td>Respondents description of decision processes associated with selection of the CTI</td>
<td>Understanding is contextually bound, vague rules; non-discriminatory; One sentence only.</td>
<td>Appears to fully understand the rationale for use</td>
<td>Appropriate identification coupled with well-founded contextual integration.</td>
<td>3</td>
</tr>
<tr>
<td>Respondents description of the interpretive steps associated with the use of the CTI</td>
<td>Understanding appears strongly linked to Procedural only; One sentence only.</td>
<td>Exhibits understanding of steps fueled by possible results and not just instrument.</td>
<td>Level of understanding is less procedural and more process oriented fueled by multiple sources.</td>
<td>2</td>
</tr>
<tr>
<td>Level of respondents description of ILP development based on CTI results</td>
<td>Vague generalities; overly tied to instrument; One sentence only.</td>
<td>Mildly decontextualized; results are on a continuum.</td>
<td>Uses client cues, instrument, and other available resources; dynamic.</td>
<td>3</td>
</tr>
<tr>
<td>Level of respondents descriptions of the specific interventions he or she employ based on CTI findings</td>
<td>Single intervention with poor treatment ideal; One sentence only.</td>
<td>Offers multiple approaches with contextual integration.</td>
<td>Eclectic philosophy governed by case by case approach.</td>
<td>3</td>
</tr>
</tbody>
</table>
SUPERVISOR RATING FORM

The following Supervisor Rating Form is divided into two sections. The first section is a brief questionnaire. The second section is a global rating scale. This is a single score estimation of the competency in which you perceive your supervisee uses the Career Thoughts Inventory (CTI). Criteria for the rating scale are presented in greater detail below.

Supervisee/Student Name: __________________________  Date: __________________

Section 1: Brief Questionnaire (Please approximate; review of records is not necessary)

1) Approximate the number of months you have supervised this student? ______
2) Approximate the number of CTIs you have reviewed with this supervisee? ______
3) Please check all of the following that apply to your rating for this student:
   a) Individual Supervision _____
   b) Group Supervision _____
   c) Portfolio Assessment _____
   d) Student Presentation _____
   e) other _____

Section 2: CTI Competency Rating Scale

Please rate this student as to his or her competency in the use of the Career Thoughts Inventory. Competent use is defined by a student’s appropriate selection and administration of the CTI and the appropriate interpretation and treatment plan integration of CTI results. Criteria for your single competency rating are outlined below. Please circle the one global rating that applies to the supervisee named above.

Rating guide for Supervisors: Based on the definition above, selection, administration, interpretation, and integration are key aspects of your global rating. If you believe that a student is stronger in some areas when compared to others, please use the sliding range of scores to guide your global rating. For example, if student X is a clear Novice across all four competency areas, your score will likely be a 1 or 2. If the student is proficient in selection and administration but a novice interpreter, the score might be a high Novice (rating of 3) or Low proficient (Rating of 4) depending on the relative strength of their selection and administration knowledge.

<table>
<thead>
<tr>
<th>Level</th>
<th>Definition</th>
<th>Range of Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Novice</td>
<td>Selection criteria based on memorization of CTI dimensions; uses CTI with loose rationale for use; interventions are rarely prescribed and if prescribed, only single intervention assigned.</td>
<td>1 2 3</td>
</tr>
<tr>
<td>Proficient</td>
<td>Clearly understands constructs of the CTI and the theoretical link that allows for application of results. Uses CTI results to offer multiple treatment avenues for clients; ILP steps are well integrated.</td>
<td>4 5 6</td>
</tr>
<tr>
<td>Expert</td>
<td>Understanding of CTI constructs are fully linked to the diagnostic and prescriptive qualities of the instrument. ILP steps are not procedural but well-informed by multiple sources; eclectic philosophy governed by case to case approach.</td>
<td>7 8 9</td>
</tr>
</tbody>
</table>
APPENDIX E
BACKGROUND QUESTIONNAIRE FOR PARTICIPANTS
EDUCATION AND EXPERIENCE

1. What is your CURRENT education level (choose one)?
   a. First year master’s student
   b. Second year master’s student
   c. Third year master’s student
   d. First year doctoral student with master’s
   e. Second year doctoral student with master’s
   f. Third year doctoral student with master’s
   g. Fourth year doctoral student with master’s
   h. Fifth year doctoral student with master’s
   i. First year doctoral student without master’s
   j. Second year doctoral student without master’s
   k. Third year doctoral student without master’s
   l. Fourth year doctoral student without master’s
   m. Fifth year doctoral student without master’s
   n. Master’s level clinician
   o. Doctoral level clinician

2. How many courses in assessment and test theory have you completed? ____

3. Approximately, how many CTIs have you administered? ____

4. Approximately, how many CTIs have you administered and reviewed with a supervisor? ____
Informed Consent

We wish to ensure that you understand your rights as a subject and that your consent to participate in this survey is fully informed. Your rights are listed below. Once you review this list, if you still wish to participate, please click the SURVEY button. By clicking this button, you consent to participate.

1. You have the right to terminate your participation in this research at anytime without penalty.

2. You have the right to know that the responses you offer will be used to improve graduate training models.

3. You have the right to know that the responses you offer will only be used for the purpose stated above. If any future uses are intended, the researchers will seek your approval.

4. You have the right to know that only the graduate student conducting this research will have access to your identity (i.e., name).

5. You have the right to know that your responses, once collected, will be secured via the surveymonkey website and password protected for your security. Only the graduate student conducting this research will have access to your personal identification information and responses.

6. The personal identification information obtained during this survey will remain confidential, to the extent allowed by law.

Given the above, please consent to participation by clicking the SURVEY button below. If you wish to withdraw, please click the EXIT THIS SURVEY link in the upper right corner of this form.

If you have questions about your rights as a subject/participant in this research, or if you feel you have been placed at risk, you can contact the Chair of the Human Subjects Committee, Institutional Review Board, through the Office of the Vice President for Research, at (850) 644-8633.
Consent for Supervisor Ratings

We wish to ensure that you understand your rights as a subject. Your consent is sought to allow the researchers of this study to ask your supervisor at the University Career Center to offer estimates of your competency in the use of the Career Thoughts Inventory (CTI). Regarding this request, your rights are listed below. Once you review this list, if you agree to participate, please sign and date the form below indicating that you consent to participate.

1. You have the right to terminate your participation in this research at anytime without penalty.

2. You have the right to know the following:
   a) Your supervisor will be contacted by the researchers of this study. The researchers are Roy Etheridge and Gary Peterson.
   b) Your supervisor at the Career Center will be asked to estimate your competency in the use of the Career Thoughts Inventory. This estimation will be based on a scoring criteria designed by the researchers involved in this study.
   c) Their estimations will be compared to scores derived from your survey responses from the web tool described to you in-person just prior to this request. The researchers wish to compare supervisor ratings with survey scores.
   d) Your web-based survey responses will not be shared with your supervisor. Once your supervisor provides a global rating of your competency in the use of the CTI, their participation will end.
   e) The personal identification information obtained during the survey portion of this research will remain confidential, to the extent allowed by law.
   f) You have the right to ask your supervisor about their rating.

3. You have the right to know that the responses you offer will only be used for the purpose stated above. If any future uses are intended, the researchers will seek your approval.

4. This research will be used to improve current training models.

5. If you have questions about your rights as a subject/participant in this research, or if you feel you have been placed at risk, you can contact the Chair of the Human Subjects Committee, Institutional Review Board, through the Office of the Vice President for Research, at (850) 644-8633.

Name (printed): ______________________________

Signature: ______________________________

Date: ______________

E-mail address: ______________________________
APPENDIX H
SUPERVISOR CONSENT TO PROVIDE RATINGS
Supervisor Consent to Provide Ratings

Your consent to provide estimates of the competency in which your supervisees exhibit when using the Career Thoughts Inventory (CTI) is sought. Your estimate will be one global rating based on a detailed criteria that you will receive from the researchers of this study. You will receive a Supervisor Rating Form that contains 5 brief questions, a description of the Rating Scale, and a single item rating scale. Regarding this request, your rights and responsibilities are listed below. Once you review this list, if you agree to participate, please sign and date the form below indicating that you consent to participate.

1. You have the right to terminate your participation in this research at any time without penalty.

2. You have the right to know the following:
   a) Supervisees will be informed that they have the option to approach you about the rating you provided. Researchers will not give your rating to supervisees. If supervisees request this rating from researchers, they will be instructed to contact you about your rating.
   c) Your estimations will be compared to scores derived from student survey responses. Researchers will compare supervisor ratings with survey scores.
   d) Student survey responses will not be shared with you. The level of agreement between a specific Supervisee’s scores and your ratings will not be available to you. Student personal identification information obtained during the survey portion of this research will remain confidential, to the extent allowed by law.
   e) You will have access to the total amount of agreement between supervisor ratings and cohort ratings.

3. You have the right to know that the ratings you offer will only be used for the purpose stated above. If any future uses are intended, the researchers will seek your approval.

4. This research will be used to improve current training models.

5. If you have questions about your rights as a subject/participant in this research, or if you feel you have been placed at risk, you can contact the Chair of the Human Subjects Committee, Institutional Review Board, through the Office of the Vice President for Research, at (850) 644-8633.

Supervisor Name (printed): ________________________________

Supervisor Signature: ________________________________

Date: _____________
APPENDIX I
EXPERT CONTENT RATING FORM
Matrix of Assessor Competence Content Rating Form

Rater Name: _________________________   Date: ________________

This study is being conducted to explore the feasibility of creating a competency-rating model that measures training progress in the use of assessment instruments in career counselor training. Based on 2 previous studies, 6 domains of importance have been identified. To measure the content validity of the 6 constructs selected, experts in the field of career counseling have been selected to rate the importance or theoretical relevance of each domain as it relates to assessing competence. You have been identified as an expert in this subject area. Below, there are 3 columns. The first column is a description of each domain. The second column is an example of each question used in previous and current research to stimulate subject responses. The final column is the rating column. Please circle one rating per domain based on the following criteria:

1 - No importance; Competent use of the CTI could be determined without this item.
2 – Some importance; Is useful but competent use of the CTI could be determined without this item.
3 – Important; Provides information tied to competent use of the CTI but this item is not critical.
4 – Critical; The competent use of the CTI could not be determined without this item/information.

<table>
<thead>
<tr>
<th>Construct</th>
<th>Test Question Used to Assess Competence</th>
<th>Expert Rating (Circle)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level in which the supervisee understands the purpose of the CTI.</td>
<td>From your understanding, what is the purpose of the Career Thoughts Inventory?</td>
<td>1 2 3 4</td>
</tr>
<tr>
<td>Level in which the supervisee understands the constructs of the CTI.</td>
<td>From your understanding, what constructs does the CTI measure?</td>
<td>1 2 3 4</td>
</tr>
<tr>
<td>Degree of expertise associated with supervisee decision to use instrument.</td>
<td>On what basis or circumstance, do you administer the CTI?</td>
<td>1 2 3 4</td>
</tr>
<tr>
<td>Degree of expertise the supervisee regularly exhibits when describing the interpretive steps associated with the use of the CTI.</td>
<td>What steps do you take when interpreting the results of the CTI with clients?</td>
<td>1 2 3 4</td>
</tr>
<tr>
<td>Degree of expertise the supervisee exhibits when developing an ILP based on CTI data.</td>
<td>What steps or procedures do you take when developing an ILP with CTI data?</td>
<td>1 2 3 4</td>
</tr>
<tr>
<td>Degree of expertise the supervisee exhibits when assigning specific interventions based on CTI findings.</td>
<td>What interventions do you prescribe that are linked to the CTI results?</td>
<td>1 2 3 4</td>
</tr>
</tbody>
</table>
APPENDIX J
FIELD TEST OF AMAC-CTI
Field Test of the Instrument

Purpose

To assess the trainee’s “level of cognitive operation”, a Measure of Assessor Competence was created drawing from Piaget’s cognitive theory and the competencies outlined by the APA above. Piaget’s theory of cognitive development (Piaget, 1970; Flavell, 1963; Flavell & Markman, 1983; Piaget, 1977) was used to delineate three distinct levels of performance competency. These levels included concrete, transitional, and formal operations (See Table 1).

Participants

The participants were 14 volunteer doctoral trainees in an APA-approved Counseling Psychology and School Psychology program who were enrolled in the second semester of a required practicum in a career center of a large southeastern university. Nine (9) were women, mean age = 27.4, range 23-47, 11 were Caucasian, 2 African American, and one Latino. Mean level of prior clinical/counseling experience was 2.4 years.

Instrumentation

The model is a 10 X 3 matrix with ten specific assessment competencies and three levels of mastery. The ten competencies are related to the understanding and knowledge of the instrument, when and when not to use the instrument, procedures in administration and interpretation, information shared with clients, and the prescription of interventions based on scores, whereas the three levels of mastery based are on Piagetian stage theory of cognitive development (i.e., concrete, transitional, and formal).

Procedures

Ten questions were read aloud by a trained confederate to each of the participants under standard conditions. The responses were tape-recorded and anonymously transcribed. Three raters, authors of the CTI, independently rated the responses to each question according to a 9-point scale where 1 –3 were degrees of Concrete Operations, 4 –6 Transitional Operations, and 7 –9 Formal Operations. The scores reported in the results section represent an average of all three raters.

Results

The mean rating for each response to the nine individual questions ranged from 3.0 to 5.4, SD = 1.5 to 2.4 (See Table A2). The lowest performing item related to translating scores into interventions (i.e., What interventions do you prescribe that are linked to the CTI?), whereas the
highest rated responses alluded to procedures used to interpret the results.

Table A2

*Competency Analyses*

<table>
<thead>
<tr>
<th>Item</th>
<th>Average</th>
<th>Standard Deviation</th>
<th>Cognitive Level of Attainment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Purpose</td>
<td>3.7</td>
<td>1.5</td>
<td>Transitional</td>
</tr>
<tr>
<td>Constructs</td>
<td>3.6</td>
<td>2.2</td>
<td>Transitional</td>
</tr>
<tr>
<td>Assignment</td>
<td>3.6</td>
<td>2.0</td>
<td>Transitional</td>
</tr>
<tr>
<td>Not to Assign</td>
<td>4.6</td>
<td>2.4</td>
<td>Transitional</td>
</tr>
<tr>
<td>Example Assign</td>
<td>3.9</td>
<td>2.1</td>
<td>Transitional</td>
</tr>
<tr>
<td>Example Not to Assign</td>
<td>4.5</td>
<td>2.4</td>
<td>Transitional</td>
</tr>
<tr>
<td>Interpretation</td>
<td>5.4</td>
<td>2.0</td>
<td>Transitional</td>
</tr>
<tr>
<td>Info Shared</td>
<td>4.5</td>
<td>1.7</td>
<td>Transitional</td>
</tr>
<tr>
<td>ILP Develop</td>
<td>3.0</td>
<td>1.5</td>
<td>Concrete</td>
</tr>
<tr>
<td>Example Interventions</td>
<td>4.0</td>
<td>2.2</td>
<td>Transitional</td>
</tr>
</tbody>
</table>

In terms of individuals, the mean performances across all 9 items ranged from 1.9 (mid Concrete) to 6.1 (high Transitional). Of the fourteen participants, five were rated at the Concrete level of operations while nine were at the Transitional level (See Table A3).

Table A3

*Individual Results*

<table>
<thead>
<tr>
<th>Respondent</th>
<th>Respondent Average</th>
<th>Cognitive Level of Attainment</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3.4</td>
<td>Concrete</td>
</tr>
</tbody>
</table>
Table A3 - continued

<table>
<thead>
<tr>
<th>Respondent</th>
<th>Respondent Average</th>
<th>Cognitive Level of Attainment</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>3.2</td>
<td>Concrete</td>
</tr>
<tr>
<td>3</td>
<td>3.1</td>
<td>Concrete</td>
</tr>
<tr>
<td>4</td>
<td>5.6</td>
<td>Transitional</td>
</tr>
<tr>
<td>5</td>
<td>3.6</td>
<td>Transitional</td>
</tr>
<tr>
<td>6</td>
<td>4.8</td>
<td>Transitional</td>
</tr>
<tr>
<td>7</td>
<td>4.3</td>
<td>Transitional</td>
</tr>
<tr>
<td>8</td>
<td>1.9</td>
<td>Concrete</td>
</tr>
<tr>
<td>9</td>
<td>4.2</td>
<td>Transitional</td>
</tr>
<tr>
<td>10</td>
<td>5.5</td>
<td>Transitional</td>
</tr>
<tr>
<td>11</td>
<td>2.7</td>
<td>Concrete</td>
</tr>
<tr>
<td>12</td>
<td>4.3</td>
<td>Transitional</td>
</tr>
<tr>
<td>13</td>
<td>4.6</td>
<td>Transitional</td>
</tr>
<tr>
<td>14</td>
<td>6.1</td>
<td>Transitional</td>
</tr>
<tr>
<td>Overall Avg.</td>
<td>4.1</td>
<td>Transitional</td>
</tr>
</tbody>
</table>

Alpha reliability coefficients of ratings of individual items between pairs of raters ranged from .84 to .91. The alpha reliability coefficient for the average scores for each item among the three raters was .86. The overall internal consistency of the AMAC-CTI model was .86. Item by item reliabilities ranged from .78 to .91. A review of the relative influence each item had on the overall reliability of the model indicated that one of the eleven questions negatively impacted the model’s reliability. The beta-weight associated with the question, “When do you assign the CTI?” was problematic. After carefully reviewing the coding schema, it was concluded that no error in coding was responsible for the decrease in reliability associated with this item. Upon assessing the coding methodology, it was determined that the overall impact of removing the problematic item would not jeopardize the global or specific validity of the model. Still more, the counter question of this item, “When do you not assign the CTI?” was also
removed. Upon removal of assign/not assign variables, the overall alpha improved to .87 and there was a modest but insignificant increase in average correlation (.43). The remaining item alpha weights did not negatively shift thus additional removals or consideration of empirical re-design was not necessary at this time.

Conclusions: Field Test

Based on the findings of the field test, five main conclusions were derived. First, individual performance with respect to the ten questions in the Measure of Assessor Competence and the one global score related to the use of psychological assessments can be reliably measured and evaluated in terms of a Piagetian model of cognitive development. This theoretical model is defined by the criteria in the PMTUC. Second, using AMAC-CTI questions, an individual profile of strengths and weakness can be developed to assist trainees in the development of remediation tasks for career counselors in training. Third, collective performances of trainees can be used to evaluate the effectiveness of the training program in fostering the critical skills of psychological assessment. Finally, the PMTUC model can be generalized for use with any psychological test or inventory. Of note, given the time and effort necessary to conduct a competency analysis such as this one, the costly nature of this design may lack practicality. Further development of the model will focus on methods to more efficiently measure mastery of assessment competencies. Possible methods of consideration include building an easy to follow web-based interview and conducting reliability analyses on the ten content items to shorten the interview format.
Office of the Vice President For Research
Human Subjects Committee
Tallahassee, Florida 32306-2742
(850) 644-8633 FAX (850) 644-4362

REAPPROVAL MEMORANDUM

Date: 5/9/2007

To: Roy Etheridge
2489 Sunburst Place
Tallahassee, FL 32301

Dept.: EDUCATIONAL PSYCHOLOGY AND LEARNING SYSTEMS

From: Thomas L. Jacobson, Chair

Re: Reapproval of Use of Human subjects in Research:
The Validation of a Measure of Competency in the Use of Psychological Assessment in
Career Counseling: A Piagetian Framework

Your request to continue the research project listed above involving human subjects has been approved
by the Human Subjects Committee. If your project has not been completed by 5/8/2008 please request
renewed approval.

You are reminded that a change in protocol in this project must be approved by resubmission of the
project to the Committee for approval. Also, the principal investigator must report to the Chair promptly,
and in writing, any unanticipated problems involving risks to subjects or others.

By copy of this memorandum, the Chairman of your department and/or your major professor are
reminded of their responsibility for being informed concerning research projects involving human
subjects in their department. They are advised to review the protocols of such investigations as often
as necessary to insure that the project is being conducted in compliance with our institution and with
DHHS regulations.

Cc: Gary Peterson
HSC No. 2007.307-R
REFERENCES


Thompson, A. P. (1986). Changes in counseling skills during graduate and undergraduate


BIOGRAPHICAL SKETCH

The author, Roy Etheridge, was born in Norfolk, Virginia, and raised in Virginia Beach, Virginia. He attended Kellam High School in Virginia Beach and then joined the United States Air Force in 1987. He attended military technical school at Brooks Air Force Base in San Antonio, Texas. During this training, he received a national certification as an Emergency Medical Technician. Roy served as an EMT in the Gulf War of 1991. After his honorable discharge in April 1991, he worked as a car salesman, vacuum salesman, lobbyist for an environmental protection agency, waiter, and bartender.

Roy attended American River College in Sacramento, California while on active duty and stationed at McClellan AFB (1987 to 1991). He received an A.A. with highest honors in 1990. After a series of jobs, he returned to school in 1993. He received his B.S. from Old Dominion University in Norfolk, Virginia in 1996. From 1996 to 1998, Roy attended Florida Atlantic University planning to complete a doctorate in experimental social psychology. After a change of heart, he began the Combined Program in Counseling Psychology and School Psychology at Florida State University in 1998. He married and had a son, William, in 2002 and decided to take a salaried position to support the family while his wife finished her Ph.D. She graduated in 2005 and another son, Grant, was born the same year. Roy will complete his doctoral internship under Argosy University – Schaumburg in the Illinois State Prison system in the Chicago suburbs. He expects his Ph.D. in the summer of 2008.