An Initial Study to Develop Instruments and Validate the Essential Competencies for Program Evaluators (ECPE)

A DISSERTATION
SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL
OF THE UNIVERSITY OF MINNESOTA
BY

Yuanjing Wilcox

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Jean A. King, Ph.D.

Advisor

June, 2012

© YUANJING WILCOX 2012

ACKNOWLEDGMENTS

As I reflect upon the completion of my dissertation I am grateful that so many people have helped me get to this point. First I want to thank my parents, 袁素兴 (Yuán, Sùxīng) and 薛安国 (Xuē, Ānguó), who encouraged me to seize opportunities when they are available (even when those opportunities took me to the other side of the planet) and to work hard at achieving my goals (even though there are many distractions in my new home). Next I want to thank Bob and Sharon Marcotte, whom I met in China, and who made me aware of opportunities at the University of Minnesota. The transition from China to Minnesota would have been extraordinarily difficult without their help during my first year here, but they took me in, treated me like a daughter, made me feel welcome, and helped me learn very quickly how to be independent and comfortable as a student in the United States. In addition, I would like to thank Jan Meier and Carol Allin, for their years of friendship and support.

I thank my committee, Dr. David Johnson, Dr. Richard Krueger, Dr. Laurie Stevahn, and especially Dr. Jean King. Dr. King introduced me to the field of evaluation and nurtured my interest as well as my understanding of the field. Her work obviously inspired and set the stage for my work on this dissertation, and this gave me an opportunity to make progress on an issue that I think is of great importance in the development of the field. She has been extremely generous with her time in assisting me in this project, and her insight into the many issues we discussed over that time has been invaluable.

I would also like to thank all of the participants who helped me to shape the survey questions, those who responded to the survey, and especially those who met with me for the interviews. I am very grateful to each of these people for taking time out of their busy schedules. The number and quality of the responses to the survey was great and the depth of insight I got from the interviews was crucial to my understanding of the importance of the competencies among professional evaluators.

Finally, I would like to thank my husband, Bill, who has been so helpful and understanding while his new wife spent so many hours on this enormous undertaking. His assistance has ranged from proof-reading and transcribing to encouraging, supporting, and reassuring me, and it has made this difficult process more manageable, productive, and even enjoyable.

ABSTRACT

Contemporary program evaluation emerged in the 1960s with the unprecedented Great Society programs. Over the past half of a century, program evaluation has clearly developed, yet it has not developed into a full-fledged profession. Unlike many established professions, program evaluation does not have certification or licensure for evaluators and cannot exclude unqualified persons from the profession. Evaluator competencies are one of the underpinning foundations of professional designation. Evaluator competencies demonstrate the knowledge, skills, and attitudes required in program evaluators. The Essential Competencies for Program Evaluators, developed by King et al. (2001) and revised by Stevahn et al. (2005), were one of the first efforts at specifying evaluator competencies. The purpose of this study was to develop instruments—a web-based survey and interview protocol—to initially validate the Essential Competencies for Program Evaluators using Messick's (1989, 1995a, 1995b) concept of unitary validity as a framework. The validated evaluator competencies can be used to enhance the professionalization of the field of evaluation.

Table of Contents

Acknowledgments i

Abstract iii

Table of Contents iv

List of Tables vii

List of Figures Ix

Chapter One – Introduction

Statement of the Problem 1

Purposes of the Study and Research Questions 6

Significance of the Study 7

Delimitations and Limitations 8

Definition of Terms 10

Chapter Two – Literature Review

History of Competence 13

Definition of Competence and Competency 17

Knowledge 21

Skills 22

Attitudes 23

Frameworks of Competence 23

Differential Psychology 23

Educational and Behavioral Psychology 24

Management Sciences 26

Methods for Developing Competencies 27

Existing Research on Evaluator Competencies 34

Professionalization of Evaluators 49

Summary 58

Chapter Three – Methods

Validity 62

Procedures 69

Survey Development Process 70

IRB Process 77

Survey Implementation 77

Interview Process 79

Summary 80

Chapter Four – Results

Survey Results 81

Respondents 81

Frequency Counts 83

Means and Standard Deviations 87

Interview Results 90

Interviewees 91

Comments on the ECPE 92
Certification/Credentialing 96
Findings 98
Summary 103

Chapter Five – Discussion and Conclusion

Evidence of Validity 105
Content-related Evidence 107
Substantive-related Evidence 109
Generalizability-related Evidence 110
Externally-related Evidence 112
Consequence-related Evidence 112
Implications for Future Research 114
Implications for Practice 116
Conclusion 119

References 121
Appendices
Appendix A: Essential Competencies for Program Evaluators Survey 134
Appendix B: Interview Protocols 148

List of Tables

1.1 Criteria for Judging Whether Program Evaluation Has Become a Profession 3

2.1 The Job Competence Assessment Method 31

2.2 Essential Skills Series in Evaluation 39

2.3 Essential Competencies for Program Evaluators (ECPE) 41

2.4 Knowledge Elements for Program Evaluation 44

2.5 Competencies for Canadian Evaluation Practice 45

2.6 Alignment of King et al.'s Research with Spencer and Spencer's Model 48

2.7 Definitions of Certification, Credentialing, Licensure, and Accreditation 52

3.1 Evidence Needed to Validate the Essential Competencies for Program Evaluators Using Messick's Six Criteria 68

3.2 Research Methods Used to Answer Questions about the Validity of the Essential Competencies for Program Evaluators 70

3.3 Grouping of Items for Think Aloud 74

4.1 Respondents' Self-reported Labeling of Their Experience 82

4.2 Table of the Frequency Results 84

4.3 Weighted Means and Weighted Standard Deviations 87

4.4 Summary of Expert Interview Results 96

4.5 Pros and Cons of Certification/Credentialing 97

4.6 First Ten Competences 100

4.7 Placement of Reflective Practice Competencies 102

5.1 Summary of the Evidence 107

5.2 Summary of the Conclusion 114

List of Figures

2.1	A model of Competence	20
2.2	Relationship between Competencies and Knowledge, Skills, and Attitudes	21
2.3	The Pathway of Conscious Competence Learning Matrix	26
2.4	Classic Competency Study Design	33
2.5	Important Developments in Evaluator Competencies	37
2.6	Three Pillars of Professional Designations in Canada	57
3.1	Facets of Validity	66
5.1	Weighted Mean Distribution of 61 Competencies	110

CHAPTER ONE

INTRODUCTION

Statement of the Problem

Contemporary program evaluation emerged in the 1960s with the unprecedented Great Society programs initiated by the Kennedy and Johnson administrations. Aiming to curb inequality, the U.S. government allocated billions of dollars into programs in education, health, housing, criminal justice, unemployment, and urban deterioration (Fitzpatrick, Sanders, & Worthen, 2004). To assess the effects of these federally funded programs, the Congress passed federal legislation that required the use of evaluation (Weiss, 1987; Wholey, 1986). With the passage of the Elementary and Secondary Education Act (ESEA) in 1965, massive aid was poured into education, which led to "a revolutionary expansion of educational research and evaluation" (Wholey, 1986, p. 6). The Congressional evaluation mandate created a surge of evaluations in the 1960s and 1970s.

In some cases, the surge of evaluations called upon educators, psychologists, sociologists, and economists to serve as evaluators of their own programs. Although highly trained in methodology, these new evaluators were far short of making conclusive findings (Weiss, 1987, p. 40). Together with the high-flying and unrealistic expectations from federal funders, most programs in the sixties and seventies did not seem to work as expected (Weiss, 1987, p. 41). Policy decisions to cut and keep programs were made according to ideology and anecdotal information rather than evaluative data (Weiss, 1987;

Wholey, 1986). Nevertheless, the field of program evaluation was forming. Universities started to offer courses, training, and even degree programs of evaluation. Program evaluation research organizations were founded. The number of evaluators increased. Starting in the late 1970s, the surge of program evaluation started to subside. However, there were still major federal efforts in evaluation. In addition, thousands of program evaluations were conducted at the state and local governments as well as in nonprofit organizations (Wholey, 1986, p. 7).

In over half of a century of development, program evaluation has clearly progressed, yet there are still hurdles that prevent program evaluation from being a full-fledged profession. Worthen (1994) proposed nine characteristics that a fully developed profession should possess. He compared program evaluation to these nine characteristics and identified that program evaluation possesses six out of nine characteristics. Program evaluation falls short on (1) procedures for the certification or licensure of evaluators, (2) a mechanism for the exclusion of unqualified practitioners, and (3) accreditation of preparation programs by evaluation associations. Table 1.1 shows Worthen's nine characteristics serving as criteria for judging whether program evaluation has become a profession.

Unlike many established professions (e.g., health care, teaching, counseling, criminal justice, and so on), program evaluation does not have certification or licensure for evaluators and cannot exclude unqualified persons from the practice. The certification of evaluators has been of concern for more than thirty years in the United State (Becker & Kirkhart, 1981; Worthen, 1994, 1999) and for almost two decades in Canada (Love,

1994). Most recently, the Canadian Evaluation Society (CES) has launched the Credentialed Evaluator (CE) designation in Canada as the closest alternative internationally to certification.

Table 1.1

Criteria for Judging Whether Program Evaluation Has Become a Profession

Criteria	Satisfied
1. A need for evaluators	Yes
2. Certification or licensure of evaluators	No
3. Exclusion of unqualified practitioners	No
4. Unique knowledge and skills of evaluation	Yes
5. Preparation programs for evaluators	Yes
6. Professional associations	Yes
7. Accreditation of preparation programs	No
8. Stable career opportunities	Yes
9. Standards of practice	Yes

Note. Adapted from “Is Evaluation a Mature Profession That Warrants the Preparation of Evaluation Professionals?” by B. R. Worthen, 1994, *New Directions for Program Evaluation*, No. 62, pp. 3-15.

According to the Canadian Evaluation Society (CES), a Credentialed Evaluator (CE) is a credential holder who “has provided evidence of the education and experience required by the CES to be a competent evaluator” (CES website: http://www.evaluationcanada.ca). Evaluator competencies became one of the underpinning foundations of this professional designation. Evaluator competencies demonstrate the knowledge, skills, and attitudes required of program evaluators. The Professional Designation Core Committee (PDCC) of the CES examined different sources of competencies related to evaluation (e.g., the latest version of their Essential Skills Series, their Core Body of Knowledge, the Treasury Board Secretariat Competencies for Evaluators, the Joint Committee’s Program Evaluation Standards, the

American Evaluation Association's Guiding Principles, and the United Nations Competencies for Evaluators). Subsequently, they developed the Competencies for Canadian Evaluation Practice (CCEP) based upon the Essential Competencies for Program Evaluators (ECPE) advanced by an independent group of university researchers in America (Stevahn, King, Ghere, & Minnema, 2005).

The Essential Competencies for Program Evaluators (ECPE) was one of the first efforts at specifying evaluator competencies. Although empowered by a "can-do attitude" (Stevahn, King, Ghere, & Minnema, 2005, p. 43), program evaluators, who came from fields, such as education, psychology, sociology, and economics, realized that their practice was, in some cases, more research- than evaluation-oriented. Moreover, the diverse philosophical and practical approaches made it difficult for the evaluation community to reach an agreement on a set of evaluator competencies (M. F. Smith, 1999; Worthen, 1999). King et al. (2001) clearly noticed these challenges, but they also recognized the need for evaluator competencies:

> First, because there is no standardized licensing or credentialing, anyone can claim to be an evaluator. Because of this, incompetent evaluators, charlatans, and crooks may well pass as seasoned professionals. Second, program directors who set out to hire an evaluator have no easy way of knowing who is qualified to do the job, and the reminder of "caveat emptor" may provide little comfort. Third, individuals interested in becoming evaluators may struggle to determine what they need to learn or where they should study. Fourth, those who provide professional

> development or university programs may base their curricula only on perceived needs or personal preferences. Fifth, a broader concern is the continuing lament that the field lacks program evaluation research aimed at developing and validating theory-based descriptive models to guide effective practice. (Stevahn, King, Ghere, & Minnema, 2005, pp. 44-45)

King et al. (2001) developed the taxonomy of essential competencies for program evaluators. The Essential Evaluator Competencies outlined the knowledge, skills, and attitudes an evaluator needs to have in order to conduct effective evaluation practice. In the development of the Essential Evaluator Competencies, a face validation study was performed with 31 evaluators from the Twin Cities area in Minnesota using a Multi-Attribute Consensus Reaching (MACR) procedure (King, Stevahn, Ghere, & Minnema, 2001). The same group of researchers (Stevahn, King, Ghere, & Minnema, 2005) later cross-referenced the proposed evaluator competencies to the *Program Evaluation Standards* (Joint Committee on Standards for Educational Evaluation, 1994), the *Guiding Principles* (American Evaluation Association Task Force on Guiding Principles for Evaluators, 1995), and the *Essential Skills Series* (Canadian Evaluation Society, 1999).

As a result of the studies, the Essential Competencies for Program Evaluators (ECPE) were established, including 61 items in six categories: professional practice, systematic inquiry, situational analysis, project management, reflective practice, and interpersonal competence. These competencies are intended to improve training, enhance reflective practice, promote evaluation research, and continue professionalization of the

field of evaluation (Stevahn, King, Ghere, & Minnema, 2005). A logical next step is to conduct empirical research on the ECPE.

Empirical research on competencies started with job analysis in the 1940s and 1950s in the field of management. Job analysis is a set of procedures designed to identify and describe those aspects of performance that differentiate high performers from low performers (Anastasi & Urbina, 1997). Spencer and Spencer (1993) reviewed 286 studies of entrepreneurial, technical and professional, sales, government, military, health care, education, and religious organizations and suggested three alternative methods for the design of competency studies.

As one of the most popular study designs, the classic competency study design includes six steps: (1) define performance effectiveness criteria; (2) identify a criterion sample; (3) collect data; (4) analyze data and develop a competency model; (5) validate the competency model; and (6) prepare applications of the competency model. Although the King et al. (2001) study was conducted independently, the study matched the steps of the classic competency study design. However, one step remains: the Essential Competencies for Program Evaluators (ECPE) need broader validation. This study will extend validation of the proposed evaluator competencies developed by King et al. (2001) and revised by Stevahn et al. (2005).

Purposes of the Study and Research Questions

Not only did the King et al. (2001) and Stevahn et al. (2005) studies provide the first step toward a possible agreement on evaluator competencies, but also the proposed

evaluator competencies are being used widely in the evaluation community. A question arises: Are the proposed evaluator competencies valid across the entire evaluation field? The purpose of this study is to develop instruments and initially validate the Essential Competencies for Program Evaluators (ECPE) so that the proposed competencies can be used to enhance the professionalization of the field of evaluation. Accordingly, the research questions of this study are:

1. To what extent are the Essential Competencies for Program Evaluators valid across the field of evaluation?
2. What additions, deletions, or other changes would make the Essential Competencies for Program Evaluators more representative across the field of evaluation?

Significance of the Study

This study will provide empirical evidence that speaks to the validity of the proposed Essential Competencies for Program Evaluators (ECPE) advanced by King et al. (2001) and revised by Stevahn et al. (2005). It is an integral part of the competency study design. With a list of validated evaluator competencies, there are three immediate benefits to the evaluation community. First, the validated evaluator competencies can serve as a resource for evaluation training programs. Whether the need is to structure a degree program at universities or to design a stand-alone workshop on evaluation knowledge, skills, and attitudes, the validated evaluator competencies are the starting point. Second, the validated evaluator competencies can serve as a professional

development tool for evaluators to self-reflect on their strengths and weaknesses in their day-to-day practices (Ghere, King, Stevahn, & Minnema, 2006). Third, the validated evaluator competencies can serve as the foundation for the eventual licensure or credentialing of evaluators and accreditation of training programs. Therefore, the study will help to promote the professionalization of the field of evaluation.

In addition, by examining the validity of evaluator competencies, this study also aims to advance research on evaluation. In general, empirical research has not been a focus within the field of evaluation (Christie, 2003; Shadish, Cook, & Leviton, 1991). Cousins and Earl (1999) suggest: "We need to move beyond the relatively weak connected theoretical musings and anecdotal reports of practice. . . We need to add to the empirical knowledge base through carefully developed and executed studies that have the potential to extend our theories and guide our practice, studies that manifestly strengthen the link between theory and practice" (p. 316). This study responds to this call for conducting empirical research.

Delimitations and Limitations

The purpose of this study is to develop instruments and initially validate the Essential Competencies for Program Evaluators (ECPE) developed by King et al. (2001) and revised by Stevahn et al. (2005). This study developed a web-based survey and interview protocol to collect evidence of validity of the ECPE using Messick's (1989, 1995a, 1995b) concept of unitary validity as a framework.

Participation in this study was delimited to practicing evaluators, clients who use evaluation, and students who study program evaluation and/or research methodology. Almost all participants were on the listserv of the Minnesota Evaluation Association (MN EA). They were not necessarily members of the MN EA. A small number of participants were also invited to participate in the study because of their extensive experience.

The vast majority of the participants of this study were from Minnesota. In addition, purposeful sampling was used to recruit interview participants. The sample was relatively small. As a result, this study has limited generalizability. Participants in this study may not reflect evaluators in other states or the national population.

The web-based survey consisted of a rating scale that asks participants to indicate the extent to which they believe that program evaluators need to be able to demonstrate a certain competency, whether or not they do so regularly. There was a potential limitation of such self-report data. Furthermore, the survey asked participants not to think about their own experience, but to think about evaluators in general. There was a potential limitation that evaluators may not have responded appropriately. It is possible that they would respond that what they do is "necessary" and what they do not do is "not necessary." In addition, whether or not something was perceived as necessary might not mean that it was important. This study focused on necessity, not importance.

Despite the limitation, this study employed a robust process of instrument development. The response rate of this web-based survey was close to 50 percent, which was beyond adequate for most online surveys. This study was exploratory in nature and set an example for conducting empirical research in the field of evaluation.

Definition of Terms

Accreditation

Accreditation is a mechanism whereby the educational program of an agency or educational institution is examined, by an external panel against established criteria for programs. The program, if it passes review, receives a formal document indicating that it is accredited.

Certification

Certification is a process by which a person masters certain skills and competencies in a field as assessed by an external body (usually a professional society in the area of consideration).

Competence

Competence is an abstract construct. As widely understood in medical fields, competence "is the habitual and judicious use of communication, knowledge, technical skills, clinical reasoning, emotions, values, and reflection in daily practice for the benefit of the individual and community being served" (Epstein & Hundert, 2002, p. 226).

Competency

In contrast to competence, competency, with its plural form, competencies, is a narrower, more atomistic concept to label particular abilities (M. K. Smith, 1996). In other words, a competency is a single skill or ability, which can be derived from functional job analysis.

Construct Validity

Construct validation is involved whenever a test is to be interpreted as a measure of some attribute or quality which is not "operationally defined" (Cronbach and Meehl, 1955, p. 282). It must be investigated whenever no criterion or universe of content is accepted as entirely adequate to define the quality to be measured.

Content Validity

Content validity is "an assessment of whether a test contains appropriate content and requires that appropriate processes be applied to that content" (Thorndike, 2005, p. 147). It is deductive and established by showing that the test items are "a sample of a universe in which the investigator is interested" (Cronbach & Meehl, 1955, p. 282).

Credentialing

A set of courses or other experiences a person must go through to receive a credential. Sometimes, it may be done by a professional society or by trainers as in a credential for having been trained.

Criterion Validity

Criterion validity is the degree to which test scores correlate with some chosen criterion measures of job or academic success. "The higher the correlation, the more effective the test is as a predictor and the higher is its criterion-related validity" (Thorndike, 2005, p. 157). If the test is given before the criterion is obtained, the validity is called *predictive validity*; if the test score and criterion score are determined at the same time, the validity is called *concurrent validity* (Cronbach & Meehl, 1955, p. 282).

Licensure

Licenses are awarded by states, branches of government, and legal jurisdictions. In certain profession (e.g., teachers), one must have a license to perform services or incur penalties if they are performed without a license. Many times the criteria for licensing are the same as certification and are determined by professional societies/groups.

CHAPTER TWO

LITERATURE REVIEW

This literature review will investigate the history and definition of competence and competencies in many fields in general and in the field of evaluation in particular. The concept of competence is built on three schools of thought, and various methods for developing competencies are reviewed. In addition, this review will summarize the development of evaluator certification in relation to evaluator competencies.

History of Competence

The concept of competence can be traced as far back as 3000 years ago when the Chinese employed written civil service exams, replacing recommendations by superiors, in selection for government jobs (Hoge, Tondora, & Marrelli, 2005). In the medieval age, apprenticeship was introduced. Apprentices were expected to learn skills by working with a master and were awarded credentials after they reached the standards of workmanship set by the trade (Horton, 2000).

With the industrial revolution, major socioeconomic changes took place in sectors such as agriculture, manufacturing, production, and transportation. The study of work and jobs in these sectors and skills needed to do these jobs emerged (Horton, 2000). Near the turn of the 20th century, social efficiency became a dominant social idea in the United States. Frederick Winslow Taylor, who was called the father of scientific management, became a significant figure with the development of management thinking and practice.

Taylor (1911) proposed greater division of labor, with jobs being simplified, an extension of managerial control over all elements of the workplace, and cost accounting based on systematic time-and-motion study. All of these elements were associated with the rise of the concept of competence.

The concept of competence continued to develop in management with a focus on work and employee selection (Hoge, Tondora, & Marrelli, 2005). In the 1930s in the United States, the Roosevelt administration promoted functional analysis of jobs, which resulted in the publication of a dictionary of occupational titles identifying knowledge and skills connected to different occupations (Horton, 2000). In the 1940s and 1950s, researchers started systematically identifying and analyzing broad performance factors (see Flanagan, 1954; Fleishman, 1953). Beginning in the 1960s, many psychologists researched individual variables that would effectively predict job performance without inherent bias against subgroups (Shippmann et al., 2000).

In 1978, the U.S. government published the *Uniform Guidelines on Employee Selection Procedures*. The *Guidelines* clearly stated that the selection of workers had to be based on job-related qualifications that resulted from the analysis of the essential work behaviors and desired outcomes of the job (Equal Employment Opportunity Commission, Civil Service Commission, Department of Labor, & Department of Justice, 1978; Hoge, Tondora, & Marrelli, 2005; Shippmann et al., 2000). In 1994, Congress created the National Skill Standards Board (NSSB), which became the National Skill Standards Board Institute (NSSBI) in 2003, to develop a voluntary national system of skill standards, assessment, and certification to enhance the ability of the United States to

compete effectively in the global economy (Horton, 2000; "National Skill Standards Act of 1994," 1994). Edie West, the executive director of the NSSB, summarized the standards as follows:

> Those skill standards identify what people need to know and be able to do in order to successfully perform work-related functions within an industry sector. Specifically, standards define the work to be performed, how well the work must be done, and the level of knowledge and skill required. (McCain, 2002, para. 3)

Not only has the competence movement shaped the industrial and business sectors in terms of efficiency, the movement has also reformed the field of education with the advent of competency-based education, which was linked to industrial and business models centered on specification of outcomes in behavioral objectives form (Tuxworth, 1989).

In the 1910s and 1920s, there were two prevailing and divergent approaches to curriculum design—one was the behavioristic or functional approach, represented by John Franklin Bobbitt, a well-known efficiency-minded educator, who recognized the differences among individuals and sought to “educate the individual according to his capabilities” (1912, p. 269); the other was the humanistic approach, represented by John Dewey (Kliebart, 1995; Neumann, 1979). The behavioristic approach builds curriculum according to highly refined and specifically identified skills or functions; the humanistic approach builds curriculum based on a combination of culture, personality, and

citizenship (Neumann, 1979). Competency-based education was a result of the behavioristic approach.

It is widely understood that competency-based education originated in teacher education (Hoffmann, 1999; Neumann, 1979; Tuxworth, 1989). This form of teacher education was introduced during the First World War when educators were called upon to train skilled tradesmen and technicians into instructors for inexperienced workers in the military (Neumann, 1979). Charles Allen became a renowned vocational educator. Allen (1919) adopted Taylor's work on efficiency and connected job analysis with education:

> Analyzing the trade simply means listing out all the things that the learner must be taught if he is to be taught the complete trade. If the trade is that of a carpenter, the instructor notes down all the different jobs that a carpenter has to do. If it is plumbing, or book binding, or machine shop work, the same listing of jobs must be carried out. If, in addition to the jobs themselves, there are certain special words (technical terms) whose use he must know, or special tools whose names he must be able to remember, or constructions or computations which he must be able to make, or special safety precautions that he must take, these must also be listed completely out. (p. 43)

Competency-based teacher education brought curriculum reform in the mid and late 1960s. In 1968, the U.S. Office of Education funded ten studies to develop model training programs for the preparation of elementary school teachers, all of which focused

on teacher competencies by implementing planning and monitoring systems (Johnson, Jr., 1984; Tuxworth, 1989). Eventually teacher accreditation systems were established throughout the county with minimum standards of performance and minimum levels of competence (Horton, 2000).

The impact of competency-based education and training reached various professions in the United States. For example, the health related professions, such as doctors, nurses, and care assistants, became well-known for adopting competency-based notions for both initial training and continuing professional development (Tuxworth, 1989). In addition, occupational certification and licensing began to emerge with the development of competencies in many professions, using sophisticated tests assessing knowledge directly related to specified competencies (Tuxworth, 1989). Furthermore, there have even been studies done to identify “competencies for life” for each human being. The Organization for Economic Co-operation and Development (OECD) has put forth competencies “necessary for individuals to lead a successful and responsible life” in the project *Definition and Selection of Competencies: Theoretical and Conceptual Foundations* (DeSeCo) (Salganik, 2001, p. 18).

Definition of Competence and Competency

It is necessary to make clear what competence and competency mean in the discussion of the professionalization of evaluators. As defined earlier, competence is an abstract construct. It describes the quality of being competent. It is the “habitual and judicious use of communication, knowledge, technical skills, clinical reasoning, emotions,

values and reflection in daily practice for the benefit of the individual and community being served" (Epstein & Hundert, 2002, p. 226). In contrast, competency is a more concrete concept to include particular knowledge, a single skill or ability, and attitudes. It speaks to the quality of being adequately or well qualified, physically and intellectually.

There is misuse and/or cross-use of the terms competence and competency. For example, competence is used as a synonym for performance, a skill, or personality trait (Bassellier, Reich, & Benbasat, 2001). Not only is competence treated as performance, but it is also used indiscriminately to refer to either the observable performance or the underlying neural functions that support the observable behavior (Connell, Sheridan, & Gardner, 2003). Schwandt (2008) drew attention to the different nature of competence and competency. He pointed out that to identify professional competence is "another" task than to identify competencies (Schwandt, 2008). In addition, Epstein and Hundert (2002) and Eraut (1994) agreed that professional competence is more than a demonstration of isolated competencies.

Although there is no agreed-upon definition of competence, there are researchers and organizations that have sought to depict it. A review of literature finds that competence is often associated with knowledge, skills, or attitudes that enable one to effectively perform the activities of a given occupation or function to the standards expected by someone (Bassellier, Reich, & Benbasat, 2001; Bird & Osland, 2004; Connell, Sheridan, & Gardner, 2003; OECD, 2002; Roe, 2002; Salganik, 2001; Spencer & Spencer, 1993; Weinert, 2001).

A concrete example of a definition of competence can be found in Roe's presentation about designing a competency-based curriculum for European work and organizational psychology. According to Roe (2002), people select their fields based upon their abilities, personality traits, and biographical characteristics, all of which are considered to be dispositions. The competencies[1] for a work and organizational psychologist are identifying client needs, individual assessment, work place analysis, selection system design, and organizational development. To obtain the above-mentioned competencies, Roe (2002) argues that one has to possess knowledge (e.g., cognitive theory, personality theory, performance theory, career theory, and job design theory), skills (e.g., problem analysis skills, observation skills, oral communication skills, writing skills, and team collaboration skills), and attitudes (e.g., respect for criticism, openness to criticism, involvement, customer orientation, and integrity). To summarize, competence builds on knowledge, skills, and attitudes, which are acquired through basic academic education and initial professional training and refined through professional practice. Figure 2.1 shows a model of competence based upon Roe's definition (2002; 2004).

[1] Roe used "competence" and "competences" in his article, which I believe he meant as "competence" and "competencies."

<table>
<tr><td>Professional practice →</td><td colspan="3">Competence</td></tr>
<tr><td>Initial professional training →</td><td rowspan="2">Knowledge</td><td rowspan="2">Skills</td><td rowspan="2">Attitudes</td></tr>
<tr><td>Basic academic education →</td></tr>
<tr><td>Selection →</td><td colspan="3">Dispositions</td></tr>
</table>

Figure 2.1. A model of competence.

Adapted from "Towards a Competence Based Curriculum for European W&O Psychology," by R. A. Roe. Paper presented at the Annual European Network of Work & Organizational Psychologists (ENOP) Symposia, Paris, 2004, March 26.

Applying this definition, knowledge, skills, and attitudes can be used to construct competencies. Nonetheless, only the applied knowledge, manifestation of skills that produce success, and observable behaviors related to attitudes are necessary and sufficient conditions to competencies (Schoonover Associates, 2003). For example, in a business setting, "to understand market pricing dynamics" is knowledge, while "to use understanding of market pricing dynamics to develop pricing models" is a competency. "To set up new project introduction" is a skill, while "to position a new project introduction so that it is clearly differentiated in the market" is a competency. Likewise, "wanting to do an excellent job" is an attitude, while "meeting all commitments in a timely manner" is a competency. Figure 2.2 shows the relationship among sample competencies and the knowledge, skills, and attitudes, on which they depend.

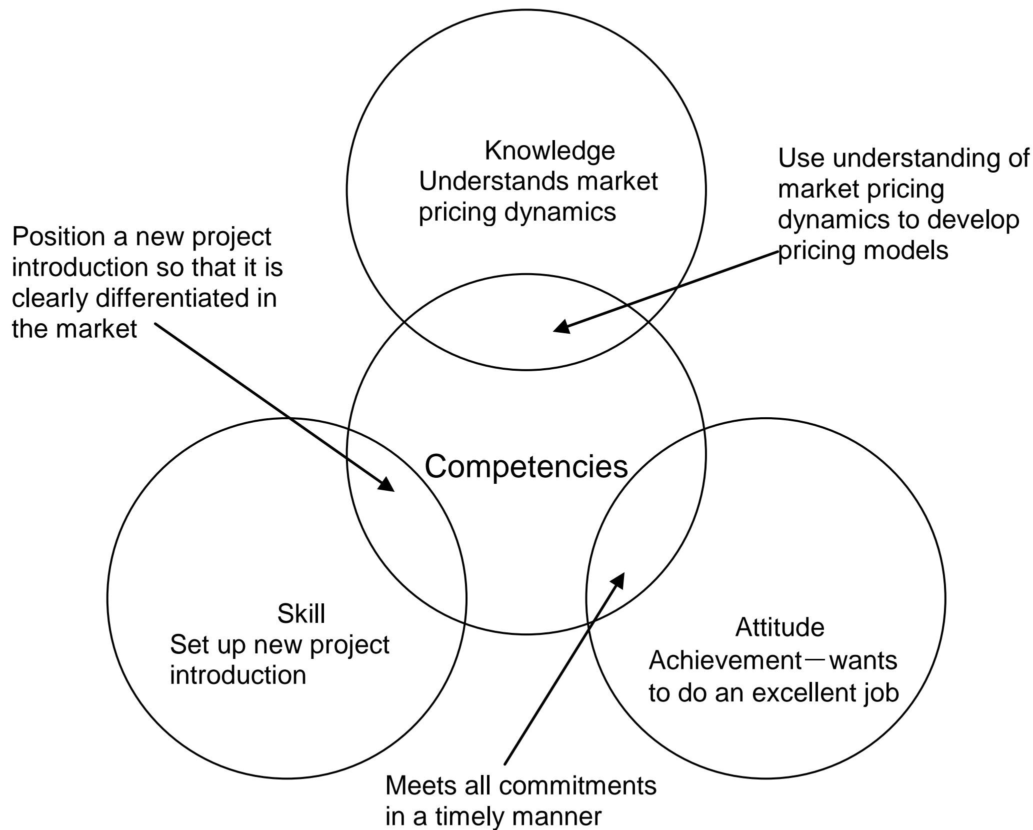

Figure 2.2. Relationship between competencies and knowledge, skills, and attitudes.

Adapted from "How do Competencies Differ from Skills and Knowledge?" by Schoonover Associates. Retrieved February 18, 2008, from http://www.schoonover.com/competency_faqs.htm

Knowledge

Knowledge is factual. It is awareness, information, or understanding about facts, rules, principles, concepts, or processes necessary to perform the tasks of a job (Hoge, Tondora, & Marrelli, 2005). Knowledge is generally acquired through formal education, on-the-job-training, and work experience (Green, 1999). Knowledge can be broken down into general knowledge and disciplinary knowledge (Evers, Rush, & Berdrow, 1998). "Any object takes space" is an example of general knowledge. Computer science and

statistics are examples of disciplinary knowledge, which is specific. To be competent in one subject, one is required to show mastery of certain specific or disciplinary knowledge in that subject.

Polanyi (1967) introduced two other types of knowledge: explicit and tacit. According to Polanyi, explicit knowledge is the knowledge that can be clearly passed on using systematic language. However, explicit knowledge is not sufficient, alone, to describe one's competence (Bassellier, Reich, & Benbasat, 2001). For example, being able to understand the techniques of basic strokes in a swimming class does not make one a good swimmer. One needs to be able to apply these techniques to be competent in swimming. The ability to perform well is tacit knowledge. Tacit knowledge is "know-how," which enables the individual to build competence by modifying his or her action based upon the results of previous actions (Leonard & Sensiper, 1998).

Skills

Boyatzis (1982) proposes that "skill is the ability to demonstrate a system and sequence of behavior that are functionally related to attaining a performance goal…it must result in something observable, something that someone in the person's environment can 'see.' For example, planning ability is a skill" (p. 33). The skill can be acquired in an educational environment or learned on the job in an informal way. Skills are often discipline-specific. In the development of competencies for a particular job or profession, specific skills of that job or profession need to be identified. To be competent, the individual needs to have minimum skills to do an effective job. Skills are not possessed in isolation. They are associated with knowledge and values and each other,

and they reinforce one another. Skills develop sequentially. Basic skills must be secured before more advanced skills (Evers, Rush, & Berdrow, 1998).

Attitudes

Knowledge and skills are considered as "surface" competencies, as they can be relatively easily identified and developed through training (Spencer & Spencer, 1993). There are, however, cognitive capacities that enable the individual to learn and perform effectively. Such capacities, including analytical thinking abilities, motivations, involvement, self-image, values, interests and preferences, are collectively considered attitudes. Attitudes are of "bi-directional causality": positive attitudes can lead to better performance, and negative attitudes can sabotage the job (Roe, 1999). Therefore, attitudes are more complex than knowledge and skills. In addition, attitudes are time-consuming to develop in competence building.

Frameworks of Competence

In the course of the evolution of the concept of competence, three main schools of competence research and practice have become influential. They are the differential-psychology approach, the educational-and-behavioral psychology approach, and the management-sciences approach (McLagan, 1997). These three approaches serve as frameworks for the research and application of the concept of competence.

Differential Psychology

"This approach focuses on human differences, especially abilities that are hard to develop...[T]hey emphasize intelligence, cognitive and physical abilities, values,

personality traits, motives, interests, and emotional qualities" (McLagan, 1997, p. 46). Under the differential-psychology framework, competencies are defined as underlying characteristics of people that are causally related to effective or superior performance in a job (Boyatzis, 1982). The use of this definition requires "inputs of individuals" in order for them to achieve competent performances (Hoffmann, 1999). The "inputs of individuals" are "underlying characteristics of people" as described in Boyatzis's definition and include people's intelligence, personality traits, and knowledge and skills that they use.

A common application of competencies with the differential-psychology framework is to single out qualities that superior performances have (McLagan, 1997). Furthermore, these competencies can be used to identify employees with high potential for leadership positions (Hoge, Tondora, & Marrelli, 2005).

Educational and Behavioral Psychology

The educational-and-behavioral psychology framework focuses on specifying the full range of competencies required for successful job performance. It emphasizes developing people so that they can be successful (Hoge, Tondora, & Marrelli, 2005). People who practice the educational-and-behavioral approach tend to well understand the "conscious competence learning matrix,"[2] which is used to explain the four psychological stages of learning new skills and techniques. Howell (1982) describes these four stages:

---

[2] The earliest origins of the conscious competence learning model are not clear, although the US Gordon Training International organization played a major role in defining it and promoting its use. From http://www.businessballs.com

> Unconscious incompetence – this is the stage where you are not even aware that you do not have a particular competence. Conscious incompetence – this is when you know that you want to learn how to do something but you are incompetent at doing it. Conscious competence – this is when you can achieve this particular task but you are very conscious about everything you do. Unconscious competence – this is when you finally master it and you do not even think about what you have such as when you have learned to ride a bike very successfully. (pp. 29-33)

Important applications of competencies in the educational-and-behavioral psychology framework are identifying competencies that employees need to become effective workers and creating performance management, training, and other development programs to help them move along the pathway from “unconscious incompetence” to “unconscious competence.” Figure 2.3 shows the pathway of the conscious competence learning matrix.

Conscious Competence Learning Matrix	Competence	Incompetence
Conscious	↓ 3. We know we know.	← 2. We know we don't know.
Unconscious	↓ 4. We don't know we know.	↑ 1. We don't know we don't know.

Figure 2.3. The pathway of conscious competence learning matrix.

From "Smooth Your Learning Journey with the Learning Matrix," by L. Swinton. Retrived April 1, 2008, from http://www.mftrou.com/support-files/learning-matrix.pdf

Management Sciences

The management-sciences approach places emphasis on the job rather than on the employee. The identification of competencies usually starts with job analysis and ends with a list of knowledge, skills, attitudes, and personal characteristics. As a result, job descriptions are developed as well as job evaluation criteria (McLagan, 1997). In addition, "job evaluation consultants, HRD administrators and compensation specialists, reengineering and total-quality experts, and task analysts are the major purveyors of this approach" (McLagan, 1997, p. 46). The common application of competencies using the management-science framework can be found in employee selection processes, in which competencies are identified to be included in job interviews and written tests (Hoge, Tondora, & Marrelli, 2005).

These frameworks provide three ways to perceive competence as an abstract construct. The differential-psychology approach emphasizes the abilities of individuals; the educational-and-behavioral approach emphasizes the developmental characteristics of competence; the management-sciences approach emphasizes job analysis rather than people. Extensive research has been done in the realm of management-sciences approach. The following section details methods to develop competence models in the management-science approach.

Methods for Developing Competencies

Empirical research on competencies started with job analysis in the 1940s and 1950s. Job analysis is a set of procedures designed to identify and describe those aspects of performance that differentiate high performers from low performers (Anastasi & Urbina, 1997). In the demand of innovative techniques in job analysis, Fleishman (1953) and Flanagan (1954) systematically analyzed supervisor job behavior and identified broad performance factors (Shippmann et al., 2000).

Fleishman's study (1953) focused on developing a method of describing leadership behavior so that different leadership patterns could be related to criteria of effectiveness in various working situations in which leaders function. He developed the *Supervisory Behavior Description Questionnaire*, which was based on the *Leader Behavior Description Questionnaire* by Hemphill. Hemphill's questionnaire contained 150 items that were derived from 1,800 original items. The 150 items were classified into

“nine a priori ‘dimensions’” by “expert judges,” and then reduced to two factors of “consideration” and “initiating structure” (Fleishman, 1953, pp. 1-2).

At the same time, Flanagan (1954) pioneered the classic “critical incident” technique (Shippmann et al., 2000). Under this technique, an incident meant any observable human activity which lent itself to the inferences about the person conducting the act. To be critical, the incident had to reveal itself with clear purpose to the observer and leave little doubt concerning its effects. “The critical incident technique consists of a set of procedures for collecting direct observations of human behavior in such a way as to facilitate their potential usefulness in solving practical problems and developing broad psychological principles” (Flanagan, 1954, p. 327). The technique was developed in 1941 through extensive studies in the Aviation Psychology Program of the United States Army Air Forces in World War II to develop procedures for the selection and classification of aircrews (e.g., six hundred and forty officers were interviewed, and a total of 3,029 critical incidents were obtained). It was formally given its present name, the critical incident technique, in 1947 by the American Institute for Research (Flanagan, 1954).

The critical incident technique includes five steps: (1) determination of the general aim of the activity, (2) development of plans and specifications for collecting factual incidents regarding the activity, (3) collection of the data, (4) analysis of the data, and (5) interpretation and reporting of the statement of the requirements of the activity. Rather than collecting opinions, hunches, and estimates, the essence of this technique is to obtain a record of specific behaviors that make a significant contribution to the activity. Applications of the critical incident technique can be found in measures of typical

performance (criteria), measures of proficiency (standard samples), training, selection and classification, job design and purification, operating procedures, equipment design, motivation and leadership (attitudes), counseling and psychotherapy (Flanagan, 1954).

In 1973, David McClelland published a seminal article entitled *Testing for Competence Rather than for Intelligence*. In his article, McClelland asserted that the traditional academic aptitude and knowledge content tests, as well as school grades and credentials, did not predict job performance or success in life. In addition, those intelligence tests were often biased against minorities, women, and persons from lower socioeconomic strata (McClelland, 1973; Shippmann et al., 2000; Spencer & Spencer, 1993). Instead, McClelland promoted the use of competency, which, he believed, would predict job performance and was not biased by race, sex, or socioeconomic factors (McClelland, 1973). He looked for research methods that would identify competency variables and found two important methods: the use of criterion samples and identification of operant thoughts and behaviors causally related to successful outcomes (McClelland, 1993).

McClelland developed a technique called the Behavioral Event Interview (BEI) in the course of conducting the Department of State study to help selecting junior Foreign Service Information Officers (FSIOs) (McClelland, 1993). The BEI grew out of Flanagan's critical incident method. However, the difference between the two is that Flanagan was interested in "identifying the task elements of jobs," while McClelland was interested in the "characteristics of the people who did a job well" (McClelland, 1993, p. 5). McClelland argues that in competency testing "what people think or say about their

motives or skills is not credible…Only what they actually do, in the most critical incidents they have faced, is to be believed" (Spencer & Spencer, 1993, p. 115). Therefore, the BEI method is designed to find out what people actually do rather than what they say.

Boyatzis (1982) extended the work of McClelland to managerial jobs (Shippmann et al., 2000). He reanalyzed all the available interviews previously conducted in a number of competence assessment studies by his colleagues at McBer and Company. The study involved 2,000 persons in 41 management jobs from 12 organizations. The research design was based on the Job Competence Assessment (JCA) method, which consists of five steps as shown in Table 2.1. In addition, the study also utilized a picture-story technique and a learning style inventory. As a result, a competence model of managers was developed with 21 characteristics.

Table 2.1

The Job Competence Assessment Method

Steps	Activities	Results
Identification of criterion measure	• Choose an appropriate measure of job performance • Collect data on managers	• Job performance data on managers
Job element analysis	• Generate list of characteristics perceived to lead to effective and/or superior job performance • Obtain item rating by managers • Compute weighted list of characteristics • Analyze clusters of characteristics	• A weighted list of characteristics perceived by managers to relate to superior performance • A list of the clusters into which these characteristics can be grouped
Behavioral Event Interviews	• Conduct Behavioral Event Interviews • Code interviews for characteristics or develop the code and then code the interviews • Relate the coding to job performance data	• A list of characteristics hypothesized to distinguish effective and/or superior from poor or less effective job performance • A list of validated characteristics or competencies
Tests and measures	• Choose tests and measures to assess competencies identified in prior two steps as relevant to job performance • Administer tests and measures and score them • Relate scores to job performance data	• A list of validated characteristics, or competencies, as assessed by these tests and measures
Competency model	• Integrate results from prior three steps • Statistically and theoretically determine and document causal relationships among the competencies and between the competences and job performance	• A validated competency model

Note. From "The Competent Manager: A Model for Effective Performance," by R. E. Boyatzis, 1982, New York: Wiley, p. 42.

To summarize 20 years of research on competency modeling, Spencer and Spencer (1993) reviewed findings from 286 studies of entrepreneurial, technical and professional, sales, government, military, health care, education and religious organizations and suggested three alternative methods for the design of competency studies: the classic study design using criterion samples, a short study design using expert panels, and studies of single incumbent and future jobs where there are not enough jobholders to offer samples of superior and average performance (p. 93). He detailed the six steps to conduct a classic competency study as shown in Figure 2.4.

More recently, competencies have come to the fore in the field of education. With rapid technological change, globalization, diversity, inequality of opportunities, poverty, conflicts and signs of ecological distress, the Organization for Economic Co-operation and Development (OECD) launched an extensive international project in late 1997, entitled *Definition and Selection of Competencies: Theoretical and Conceptual Foundations* (DeSeCo), to identify key competencies for an individual to lead a successful and responsible life while facing the challenges in society (Rychen, 2004). The project was led by the Swiss Federal Statistics Office and supported by the US National Center for Education Statistics and Statistics Canada (OECD, 2002). The key results of DeSeCo include the concept of competence as the "overarching frame of reference" and the three categories of key competencies: acting autonomously, using tools interactively, and functioning in socially heterogeneous groups (Rychen, 2004).

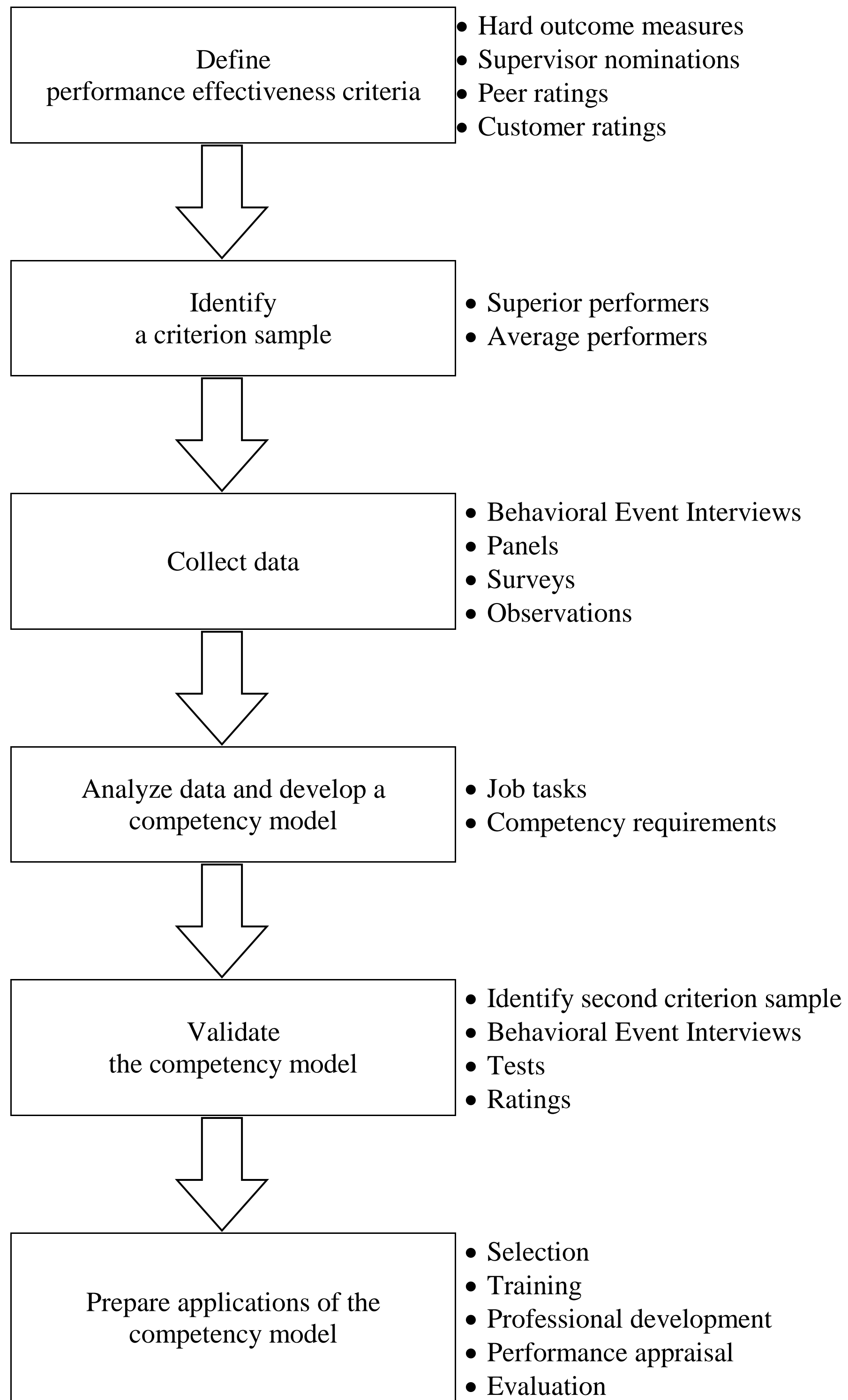

Figure 2.4. Classic competency study design.

Modified by the author from "Competence at Work: Models for Superior Performance," by L. M. Spencer & S. M. Spencer, 1993, New York: Wiley.

DeSeCo was set up as a policy-driven, theory-oriented, and multidisciplinary project, which originated in a government context in response to policy needs (Rychen, 2001). The project included a review of a number of empirical studies focusing on the development of learning outcomes (Salganik, 2001). The study also takes account of a series of discipline-oriented studies to identify theoretically grounded sets of key competencies. The perspectives on competencies and models developed by anthropologists, economists, historians, philosophers, psychologists, and sociologists were considered collectively to develop a coherent frame of reference for defining and selecting key competencies (Rychen, 2001).

Existing Research on Evaluator Competencies

As an emerging profession, the field of program evaluation has been debating the possibilities of evaluator competencies for more than two decades. On the one hand, the field of program evaluation is in need of competencies for evaluators as a foundation to guide training programs for novice practitioners, develop continuing education for experienced professionals, and conduct periodic reviews to ensure the integrity of the field (Stevahn, King, Ghere, & Minnema, 2005). On the other hand, the evaluation community has not fully reached a consensus on a list of evaluator competencies that would represent diverse philosophical and practical approaches currently used in program evaluation (King, Stevahn, Ghere, & Minnema, 2001; M. F. Smith, 1999; Worthen, 1999). For example, Kirkhart (1981, p. 189) once acknowledged the nature of program evaluation as being “flexible,” “situation-specific,” and “conceptually muddled”;

therefore, she called the attempt to draft a list of “fixed” skills for program evaluation a “folly.”

With the professionalization of program evaluation, evaluation standards were developed by the Joint Committee on Standards for Educational Evaluation (1981) in four identified key areas: utility, feasibility, propriety, and accuracy.[3] “The Standards offer a framework for understanding the skills and competencies needed to conduct program evaluations… In so doing, the Standards implicitly assume certain requisite skills, knowledge, and competencies on the part of the evaluator(s)” (Kirkhart, 1981, p. 188). In 1995, a task force of the American Evaluation Association created the *Guiding Principles for Evaluators*. “The principles are intended to guide the professional practice of evaluators, and to inform evaluation clients and the general public about the principles they can expect to be upheld by professional evaluators” (American Evaluation Association, 2004). Both the Standards and the Principles, however, “do not directly address the competencies an evaluator needs to function effectively in specific contexts” (King, Stevahn, Ghere, & Minnema, 2001, p. 230).

Smith (1999) suggests three options for developing evaluator competencies in the context of evaluator certification: (1) a self-generated list of competencies by professional evaluators, (2) a job analysis approach based on the input from evaluation recipients, and (3) identifying evaluator competencies according to the American Evaluation Association’s (AEA) *Guiding Principles for Evaluators* (2004). She clearly states that all three options have limitations. The self-generated list would be “self-

---

[3] The *Program Evaluation Standards* (3rd ed., 2011) includes metaevaluation as a fifth attribute called evaluation accountability, which is elevated from the accuracy standard.

serving," not user-sensitive, and lacking empirical basis; the job analysis approach is costly; and the *Guiding Principles* are too general to apply (M. F. Smith, 1999, pp. 525-526).

Patton (1990) and Scriven (1996), two widely respected experts in the field of program evaluation, have informally identified lists of competencies for evaluators. Upon delivering the keynote address for the Australasian Evaluation Society regarding the challenge of program evaluation being a profession, Patton proposed that a competent evaluator should possess multiple and diverse methods, communication skills, conceptualization and program logic capabilities, consulting skills, interpersonal competence, political sophistication, knowledge of how organizations work, creativity, and verbal and written presentation skills (1990, p. 48). Scriven challenged evaluators to have "reasonable competence" of being able to understand and apply basic qualitative and quantitative methodologies, validity theory, generalizability theory, meta-analysis, legal constraints on data control and access, ethical analysis, needs assessment, cost analysis, internal synthesis models and skills, conceptual geography, and evaluation-specific report design, construction, and presentation (1996, p. 160).

Kirkhart (1981) and Mertens (1994), among others, reflected on evaluation practice and organized the skills and knowledge of program evaluation into conceptual frameworks. Kirkhart identified eight major descriptive categories of evaluator competencies: (1) methodological skills, (2) knowledge areas providing substantive background, (3) systems analysis skills, (4) political savvy and understanding, (5) professional ethics, (6) management skills, (7) communication skills, and (8)

interpersonal skills or character traits (1981, pp. 188-189). Mertens divided those skills and knowledge into four categories: (1) those unique to evaluation, (2) those associated with typical training in the methodology of research and inquiry, (3) those in such related areas as political science or anthropology, and (4) those that are discipline-specific (1994, p. 19). Figure 2.5 outlines important developments in evaluator competencies.

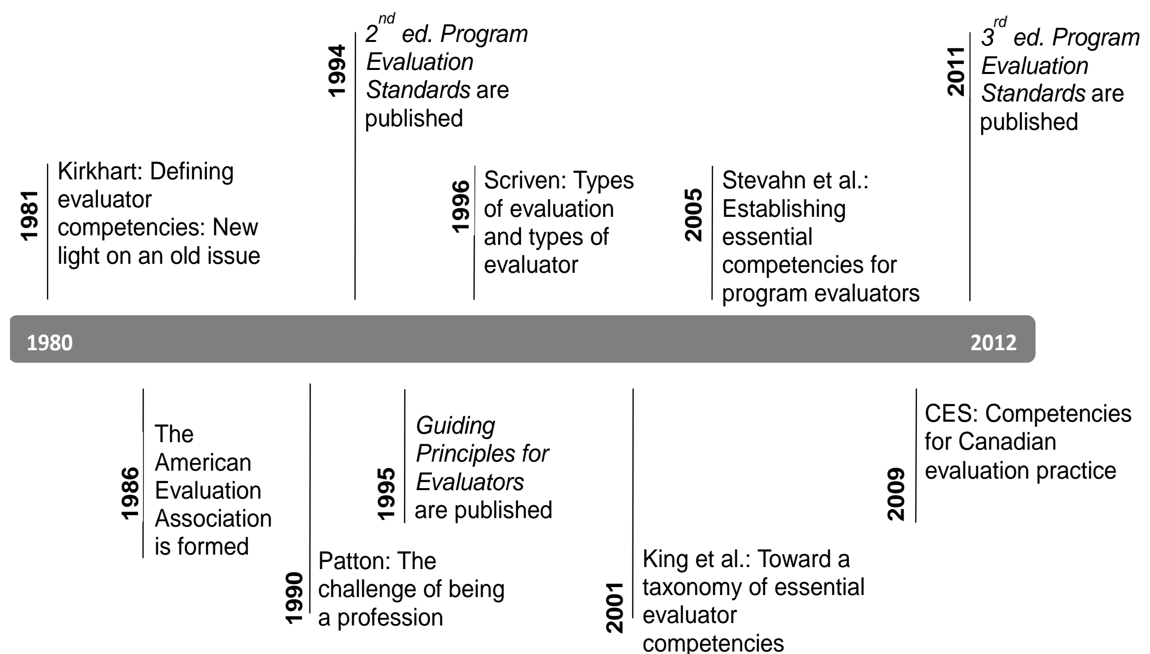

Figure 2.5. Important developments in evaluator competencies.

As a professional organization, the Canadian Evaluation Society (CES) has been in strong support of developing evaluator competencies. CES charged its Professional Development Committee to promote the professional practice of evaluation. As a result, the Essential Skills Series in Evaluation (ESS) was created in 1999. ESS, as shown in Table 2.2, is a four-day series of workshops guided by an experienced evaluator that introduces evaluation concepts and methods including: (1) understanding program

evaluation, (2) building an evaluation framework, (3) improving program performance, and (4) evaluating for results (Canadian Evaluation Society, 2004). ESS serves as an overview of essential competencies required in program evaluation.

Table 2.2

Essential Skills Series in Evaluation

Understanding program evaluation	Building an evaluation framework
• Key terms and concepts; • Benefits of program evaluation; • Basic steps in the evaluation process; • Major approaches to program evaluation; • Formatting evaluation questions; • Designing an evaluation; • Evaluating with limited resources; • Analyzing and reporting evaluation results; • Reducing resistance to evaluation; • Involving staff and clients in the evaluation process; • Increasing evaluation utilization; • Making evaluations ethical and fair.	• Identifying who the client is and what the client needs; • Basic concepts of needs assessment; • Major approaches to assessing client needs; • Evaluation methods for "getting close to the client"; • Building an evaluation framework through logic models; • Involving managers and staff in building an evaluation framework; • Relating program design to client needs; • Defining program components; • Formulating indicators for program success; • Using the evaluation framework for linking program performance to client needs.

Improving program performance	Evaluating for results
• Using evaluation as a management tool for improving program performance and enhancing internal accountability; • Basic concepts of monitoring and process evaluation; • Monitoring program performance with existing administrative data and information systems; • Developing ongoing data collection instruments and procedures; • Linking process evaluation to program decision-making; • Assessing client satisfaction; • Understanding continuous quality improvement; • Using program evaluation for building a "learning organization."	• Defining program results; • Major approaches to evaluating results; • Developing results measures; • Designing outcome evaluations; • Validity and reliability; • Appropriate use of quantitative and qualitative techniques; • Relating program results to program costs; • Understanding program benefits; • Measuring program equity and responsiveness to community needs; • Communicating evaluation findings; • Using evaluations to improve program effectiveness and accountability.

Note. Adapted from “Essential Skills Series in Evaluation” by the Canadian Evaluation Society. From http://www.evaluationcanada.ca/site.cgi?s=3&ss=31&_lang=en

Unlike the extensive research on competencies conducted in management, the research on evaluator competencies is scarce. Most available research on evaluator competencies is theoretical in nature. There is no dearth of points of view on whether or not to have evaluator competencies and what competencies an evaluator ought to have, but there has been lack of empirical studies on these issues. Although several frameworks identifying evaluator competencies have been proposed, none of these frameworks "have been derived from a systematic process or validated by empirical consensus building among diverse professionals in the field" (King, Stevahn, Ghere, & Minnema, 2001, p. 230). However, there is a silver lining in the cloud.

It was not until the new millennium that two projects were conducted to develop inventories and descriptions of competencies for evaluators (Huse & McDavid, 2006). One of the projects was sponsored by the Canadian Evaluation Society (hereafter called the Canadian project). The other was conducted independently by a group of university researchers in the United States to develop Essential Competencies for Program Evaluators (ECPE) (hereafter called the ECPE project).

The ECPC project, which was not sponsored by the American Evaluation Association, was conducted by King et al. (2001) in the USA. King et al. developed the taxonomy of essential competencies for program evaluators. The taxonomy of essential evaluator competencies outlined the capabilities an evaluator needs to have in order to conduct effective evaluation practice. The same group of researchers (Stevahn, King, Ghere, & Minnema, 2005) further refined the evaluator competencies and cross-referenced them to the *Program Evaluation Standards* (Joint Committee on Standards for

Educational Evaluation, 1994) and the *Guiding Principles* (American Evaluation Association Task Force on Guiding Principles for Evaluators, 1995) as well as the *Essential Skills Series* (Canadian Evaluation Society, 1999). In Ghere et al. (2006) the group went one step further. They turned the essential competencies into a self-assessment instrument with which evaluators could reflect on their evaluation practice.

The Essential Competencies for Program Evaluators were developed out of an exercise in an evaluation studies colloquium, a graduate level class (King, Stevahn, Ghere, & Minnema, 2001). A face validation study was then performed with 31 evaluators from the Twin Cities area in Minnesota by using a Multi-Attribute Consensus Reaching (MACR) procedure (Stevahn, King, Ghere, & Minnema, 2005). As a result of the studies, the Essential Competencies for Program Evaluators were established, including 61 items in six categories: professional practice, systematic inquiry, situational analysis, project management, reflective practice, and interpersonal competence (see Table 2.3). These competencies are intended to improve training, enhance reflective practice, promote evaluation research, and continue professionalization of the field of evaluation (Stevahn, King, Ghere, & Minnema, 2005). The ECPE project "…continues to be an active research, consultation and professional development process" (Huse & McDavid, 2006, p. 11).

Table 2.3

Essential Competencies for Program Evaluators (ECPE)

1.0 Professional Practice	
1.1	Applies professional evaluation standards
1.2	Acts ethically and strives for integrity and honesty in conducting evaluations
1.3	Conveys personal evaluation approaches and skills to potential clients

1.4 Respects clients, respondents, program participants, and other stakeholders
1.5 Considers the general and public welfare in evaluation practice
1.6 Contributes to the knowledge base of evaluation

2.0 Systematic Inquiry

2.1 Understands the knowledge base of evaluation (terms, concepts, theories, assumptions)
2.2 Knowledgeable about quantitative methods
2.3 Knowledgeable about qualitative methods
2.4 Knowledgeable about mixed methods
2.5 Conducts literature reviews
2.6 Specifies program theory
2.7 Frames evaluation questions
2.8 Develops evaluation design
2.9 Identifies data sources
2.10 Collects data
2.11 Assesses validity of data
2.12 Assesses reliability of data
2.13 Analyzes data
2.14 Interprets data
2.15 Makes judgments
2.16 Develops recommendations
2.17 Provides rationales for decisions throughout the evaluation
2.18 Reports evaluation procedures and results
2.19 Notes strengths and limitations of the evaluation
2.20 Conducts meta-evaluations

3.0. Situational Analysis

3.1 Describes the program
3.2 Determines program evaluability
3.3 Identifies the interests of relevant stakeholders
3.4 Serves the information needs of intended users
3.5 Addresses conflicts
3.6 Examines the organizational context of the evaluation
3.7 Analyzes the political considerations relevant to the evaluation
3.8 Attends to issues of evaluation use
3.9 Attends to issues of organizational change
3.10 Respects the uniqueness of the evaluation site and client
3.11 Remains open to input from others
3.12 Modifies the study as needed

4.0 Project Management

4.1 Responds to requests for proposals
4.2 Negotiates with clients before the evaluation begins
4.3 Writes formal agreements
4.4 Communicates with clients throughout the evaluation process
4.5 Budgets an evaluation

4.6	Justifies cost given information needs
4.7	Identifies needed resources for evaluation, such as information, expertise, personnel, instruments
4.8	Uses appropriate technology
4.9	Supervises others involved in conducting the evaluation
4.10	Trains others involved in conducting the evaluation
4.11	Conducts the evaluation in a nondisruptive manner
4.12	Presents work in a timely manner
5.0 Reflective Practice	
5.1	Aware of self as an evaluator (knowledge, skills, dispositions)
5.2	Reflects on personal evaluation practice (competencies and areas for growth)
5.3	Pursues professional development in evaluation
5.4	Pursues professional development in relevant content areas
5.5	Builds professional relationships to enhance evaluation practice
6.0. Interpersonal Competence	
6.1	Uses written communication skills
6.2	Uses verbal/listening communication skills
6.3	Uses negotiation skills
6.4	Uses conflict resolution skills
6.5	Facilitates constructive interpersonal interaction (teamwork, group facilitation, processing)
6.6	Demonstrates cross-cultural competence

Note. From “Establishing Essential Competencies for Program Evaluators,” by L. Stevahn, J. A. King, G. Ghere, and J. Minnema, 2005, *American Journal of Evaluation*, Vol. 26, No. 1, pp. 43-59.

The Canadian project was conducted by Zorzi, McGuire, and Perrin (2002) under the auspices of the CES. Zorzi, McGuire, and Perrin were able to broadly engage evaluators in the discussion of developing “a core body of knowledge (CBK)” for program evaluation. They conducted two internet consultations with the evaluation community, two discussion sessions with delegates at the CES 2002 National Conference, and online discussions among the members of an invited, international expert reference panel (Zorzi, McGuire, & Perrin, 2002, p. i). They identified a list of 151 knowledge elements and grouped them into 23 more general elements that were further categorized into six categories: ethics, evaluation planning and design, data collection, data analysis

and interpretation, communication and interpersonal skills, and project management (see Table 2.4) (Zorzi, McGuire, & Perrin, 2002, pp. 31-46).

Table 2.4

Knowledge Elements for Program Evaluation

Ethics
• Ethical conduct
• Competence and quality assurance
Evaluation Planning and Design
• Understanding the program
• Assessing readiness for the evaluation
• Focusing the evaluation
• Systems theory, organizational development, and change
• Specific types of evaluation
• History of evaluation, evaluation theory, and evaluation models
• Research design
• Constructing meaning
• Selecting appropriate data collection and analysis methods
• Effective practices in applied research
Data Collection
• Sampling
• Measurement issues
• Data collection methods
Data Analysis and Interpretation
• Qualitative analysis
• Quantitative analysis
• Determining merit or worth
• Critical thinking skills
Communication and Interpersonal Skills
• Interpersonal skills
• Reporting skills
• Other communication skills
Project Management
• Managing evaluation projects

Note. From "Evaluation benefits, outputs, and knowledge elements: Canadian Evaluation Society project in support of advocacy and professional development," by R. Zorzi, M. McGuire, and B. Perrin, 2002. From http://consultation.evaluationcanada.ca/pdf/ZorziCESReport.pdf

Though the CES had been seeking the “foundational knowledge” required of program evaluators by developing and updating the Essential Skills Series and commissioning the research project of the Core Body of Knowledge (CBK), it was not until more recently that the Professional Designations Core Committee (PDCC) of the CES proposed the Competencies for Canadian Evaluation Practice (See Table 2.5). After member consultation and expert validation in 2008 and 2009, the Competencies for Canadian Evaluation Practice were approved by the CES membership in May 2009.

Built on the Essential Competencies for Program Evaluators (ECPE) advanced by Stevahn, King, Ghere and Minnema (2005), the PDCC of CES conducted a crosswalk of the latest version of ESS, the CBK Study, Treasury Board Secretariat Competencies for Evaluators in the Government of Canada, the Joint Committee Program Evaluation Standards, the American Evaluation Association’s Guiding Principles for Evaluators, and the United Nations Competencies for Evaluators in the United Nations System. The crosswalk identified the gaps and overlaps among existing evaluator competencies so as to provide a comprehensive list of evaluator competencies for Canadian evaluation practice.

Table 2.5

Competencies for Canadian Evaluation Practice

1.0 Reflective Practice competencies focus on the fundamental norms and values underlying evaluation practice and awareness of one’s evaluation expertise and needs for growth.	
1.1	Applies professional evaluation standards
1.2	Acts ethically and strives for integrity and honesty
1.3	Respects all stakeholders
1.4	Considers human rights and the public welfare in evaluation practice
1.5	Provides independent and impartial perspective

1.6 Aware of self as an evaluator (knowledge, skills, dispositions) and reflects on personal evaluation practice (competencies and areas for growth)
1.7 Pursues professional networks and self development to enhance evaluation practice

2.0 Technical Practice competencies focus on the specialized aspects of evaluation, such as design, data collection, analysis, interpretation, and reporting.

2.1 Understands the knowledge base of evaluation (theories, models, types, methods and tools)
2.2 Specifies program theory
2.3 Determines the purpose for the evaluation
2.4 Determines program evaluability
2.5 Frames evaluation questions
2.6 Develops evaluation designs
2.7 Defines evaluation methods (quantitative, qualitative or mixed)
2.8 Identifies data sources
2.9 Develops reliable and valid measures/tools
2.10 Collects data
2.11 Assesses validity of data
2.12 Assesses reliability of data
2.13 Assesses trustworthiness of data
2.14 Analyzes and interprets data
2.15 Draws conclusions and makes recommendations
2.16 Reports evaluation findings and results

3.0 Situational Practice competencies focus on the application of evaluative thinking in analyzing and attending to the unique interests, issues, and contextual circumstances in which evaluation skills are being applied.

3.1 Respects the uniqueness of the site
3.2 Examines organizational, political, community and social contexts
3.3 Identifies impacted stakeholders
3.4 Identifies the interests of all stakeholders
3.5 Serves the information needs of intended users
3.6 Attends to issues of evaluation use
3.7 Attends to issues of organizational and environmental change
3.8 Applies evaluation competencies to organization and program measurement challenges
3.9 Shares evaluation expertise

4.0 Management Practice competencies focus on the process of managing a project / evaluation, such as budgeting, coordinating resources and supervising.

4.1 Defines work parameters, plans and agreements
4.2 Attends to issues of evaluation feasibility
4.3 Identifies required resources (human, financial and physical)
4.4 Monitors resources (human, financial and physical)
4.5 Coordinates and supervises others
4.6 Reports on progress and results

4.7	Identifies and mitigates problems / issues
5.0 Interpersonal Practice competencies focus on people skills, such as communication, negotiation, conflict resolution, collaboration, and diversity.	
5.1	Uses written communication skills and technologies
5.2	Uses verbal communication skills
5.3	Uses listening skills
5.4	Uses negotiation skills
5.5	Uses conflict resolution skills
5.6	Uses facilitation skills (group work)
5.7	Uses interpersonal skills (individual and teams)
5.8	Uses collaboration / partnering skills
5.9	Attends to issues of diversity and culture
5.10	Demonstrates professional credibility

Note. From “Companion Document for Competencies for Canadian Evaluation Practice,” by Canadian Evaluation Society, 2009. From http://www.evaluationcanada.ca.

Comparing the development of the Competencies for Canadian Evaluation Practice (CCEP) and the Essential Competencies for Program Evaluators (ECPE), different approaches are observed. The CCEP was derived from an extensive review of literature. The authors cross-examined existing competencies in relation to evaluation. This process had the benefit of achieving thorough findings within a relatively short period of time. However, the results were not empirically validated. Even though the Professional Designations Core Committee of the CES did perform member consultation by developing an online survey seeking CES members’ input on the proposed CCEP, only 5.5% of the CES membership, ninety-nine members in total, responded to the survey. In addition, the survey was an opinion survey. It simply asked yes/no questions without probing the reasons behind the answers. As a result, the survey was too general to gather detailed information on specific competencies.

King et al.’s research on the Essential Competencies for Program Evaluator fits Spencer and Spencer’s (1993) model of a classic competency study design. The classic

competency study design includes six steps: (1) define performance effectiveness criteria; (2) identify a criterion sample; (3) collect data; (4) analyze data and develop a competency model; (5) validate the competency model; and (6) prepare applications of the competency model. Table 2.6 shows how King et al.'s research is in the alignment with Spencer and Spencer's classic competency study design. It is noted that there is a gap between step five and step six—the Essential Competencies for Program Evaluators have been used without validation. This gap remains a research opportunity.

Table 2.6

Alignment of King et al.'s Research with Spencer and Spencer's Model

Step	Spencer and Spencer's model (1993)	King et al.'s research (2001)
1	Define performance effectiveness criteria	Class exercise in an evaluation colloquium
2	Identify a criterion sample	Thirty-one evaluators from the Twin Cities area in Minnesota
3	Collect data	Multi-Attribute Consensus Reaching Procedure (MACR)
4	Analyze data and develop a competency model	Essential Competencies for Program Evaluators (ECPE)
5	Validate the competency model	Not complete; need a second criterion sample
6	Prepare applications of the competency model	The ECPE has been used by CES to develop competencies for Canadian evaluation practice

Both the Essential Competencies for Program Evaluators and the Competencies for Canadian Evaluation Practice seek to exhibit the knowledge, skills, and attitudes needed for an evaluator to produce accurate and credible evaluations. The development of evaluator competencies may directly influence the professionalization of evaluators.

Professionalization of Evaluators

The professionalization of the field of program evaluation has undergone several milestones over the past three decades. In 1986, two organizations in the United States—the Evaluation Research Society and the Evaluation Network—merged to form the American Evaluation Association (AEA). AEA adopted the *Program Evaluation Standards*, developed the *Guiding Principles for Evaluators*, and publishes journals devoted to program evaluation (Stevahn, King, Ghere, & Minnema, 2005). Fitzpatrick, Sanders, and Worthen (2004) suggest the field of program evaluation has been steadily moving toward a profession. However, the maturation of program evaluation as a full-fledged profession has not yet been achieved (Fitzpatrick, Sanders, & Worthen, 2004; King, Stevahn, Ghere, & Minnema, 2001; Worthen, 1994). Unlike many established professions (e.g., health care, teaching, counseling, criminal justice, and so on), program evaluation does not have certification or licensure for evaluators and cannot exclude unqualified persons from the profession (Fitzpatrick, Sanders, & Worthen, 2004; Worthen, 1994).

The certification of evaluators has been of concern for more than thirty years in the United States (Becker & Kirkhart, 1981; Worthen, 1994, 1999) and for almost two decades in Canada (Love, 1994). Becker and Kirkhart (1981) proposed two "common approaches"—licensure and accreditation—to assuring the quality of the profession. They clearly stated that "…the underlying motivation for accreditation/licensure is protecting the public welfare…" (Becker & Kirkhart, 1981, p. 155).

In 1995, Sanders and Bickman submitted a memorandum to the Board of Directors of the American Evaluation Association suggesting development of a voluntary system of certification (Altschuld, 1999b). A Task Force was then formed to prepare a report on certification for the AEA Board to recommend future actions. “Activities of the Task Force included: conducting a national survey of AEA membership; reviewing the literature; submitting a report to the Board in 1997 (Altschuld, 1997); and participating in a panel discussion of and facilitating a debate on certification at the 1997 and 1998 AEA annual conferences” (Altschuld, 1999b, p. 482). A hot discussion of evaluator certification ensued.

Bickman, the former AEA president, represented a strong voice in favor of adopting a voluntary system of certification (Bickman, 1997). He believed that “…the work that is necessary to establish a clear identity of evaluation will help us as both a profession and as the premier organization of evaluators” (Bickman, 1999, p. 519). Bickman drew comparison with the American Sociological Association and the American Institute of Certified Public Accountants and concluded that the reason why accountancy is such a powerful professional group is because of the certification of its members and the accreditation of its educational programs (Bickman, 1997). Bickman emphasized that “…certification offers us a way to focus on who we are and to present a unified face to society…We need to move ahead in professionalizing evaluation or else we will just drift into oblivion” (Bickman, 1997, p. 8).

On the contrary, M. F. Smith opposed certification. She argued that AEA should not undertake certification or similar processes because “it seems unlikely to me that the

American Evaluation Association (AEA) could develop, implement, and maintain an effective process, [and] even if it could, AEA does not have the funds to do so" (M. F. Smith, 1999, p. 521). One of the strong points for Smith to oppose certification was the results from the survey of AEA members' opinions concerning voluntary evaluator certification (Jones & Worthen, 1999). The survey results revealed that at least 28% of the 174 respondents specifically did not favor certification (Jones & Worthen, 1999), and 67% of them did not have confidence that a certification process would be effective (M. F. Smith, 1999).

Considering the complication of the certification process, Worthen (1999) identified four major challenges that confront any effort to develop a viable evaluator certification system. These challenges are:

> (1) determining what basic approach to certification should be taken and what type of evidence is most compelling; (2) reaching agreement on what evaluation is and what core knowledge and skills all evaluators should possess; (3) constructing professionally and legally defensible certification procedures and instruments; and (4) gathering support for a mandatory certification process. (Worthen, 1999, p. 545)

To determine a basic approach to certification, Worthen (1999, pp. 545-546) proposed four possible approaches: certification based on formal training, certification based on evaluation experience, certification based on performance, and certification based on competencies. He considered the competency-based

approach as the most promising approach due to its feasibility, objectivity, cost-effectiveness, and applicability (Worthen, 1999).

Consistently endorsing professionalization of evaluators, Altschuld has modified his position over time from "pro" certification to advocating a voluntary system for credentialing evaluators due to the complexity of the issue (Altschuld, 1999a). Altschuld acknowledges that credentialing is "a partial first step toward establishing some boundaries for the field of evaluation" (1999a, p. 510). He defines credentialing as "the fact that a person has studied a field and completed specified courses and activities in that field" (Altschuld, 1999a, pp. 507-508), which is different from certification that indicates "the individual has attained a certain level of knowledge and skills in a field, usually as determined through an examination or a series of examinations" (Altschuld, 1999a, p. 508). To clarify the confusion among terms, such as licensure, credential, certification, and accreditation, Altschuld (2005) provides definitions and descriptions for each of them (see Table 2.7).

Table 2.7

Definitions of Certification, Credentialing, Licensure, and Accreditation

Terms	Definition	Description
Certification	A process by which a person masters certain skills and competencies in a field as assessed by an external body (usually a professional society in the area of consideration).	• Most often done through a formal test or set of tests (certification exams) as in law, medicine, engineering, etc. • Certifying body may be legally liable for the skills that they designate as being attained by an individual • Certification may have to be periodically renewed most frequently (but not always) via continuing education

Terms	Definition	Description
Credentialing	A set of courses or other experiences a person must go through to receive a credential. May be done by a professional society or sometimes by trainers as in a credential for having been trained.	• Does not specify the skill set attained by the person credentialed, only that they have gone through delineated experiences and courses • Tests or certification exams may be, but generally are not, used for credentialing; instead it is the courses or training experiences that the individual has taken • The legal implications for credentialing are less than for certification
Licensure	Licenses are awarded by states, branches of government, and legal jurisdictions. One must have a license to perform services or undergo penalties if they are performed without a license. Many times the criteria for licensing are the same as certification and are determined by professional societies/groups.	• One may be certified but not licensed as in the case of a physician who has passed the necessary medical examinations but is found to have defrauded patients or illegally used drugs • Legal jurisdictions set up review panels in cases where there is malfeasance or unsafe practice • Control of licensure resides outside of the professional group but is almost always highly influenced by it
Accreditation	A mechanism whereby the educational program of an agency or educational institution is examined, by an external panel against established criteria for programs. The program, if it passes review, receives a formal document indicating that it is accredited.	• Accreditation is for a program whereas certification, credentialing, and licensure relate to an individual • Accreditation reviews rely on the courses and experiences that comprise a program, the skills gained by those going through it, their proficiencies as determined by tests and other outcome measures, and the processes through which the program is delivered

Note. From "Certification, Credentialing, Licensure, Competencies, and the Like: Issues Confronting the Field of Evaluation," by J. W. Altschuld, *Canadian Journal of Program Evaluation*, 20(2), 157-168.

On the journey toward professionalization, the Canadian Evaluation Society (CES) has traveled farther than its American counterpart. At about the same time that the American Evaluation Association (AEA) hotly debated the issue of certification in the 1990s, so did the CES. At that time, the CES commissioned a fact-finding project on

professional designations and certification in other professional organizations (Long & Kishchuk, 1997) and decided not to move ahead with professional designation. However, the position of the CES later changed, although AEA remains highly divided on the issue (Altschuld, 2007).

The increasing interest in a system of professional designation among the Canadian evaluation community was reflected in a survey conducted by Borys, Gauthier, Kishchuk, and Roy (2005). For example, of the 1,005 respondents, 64% indicated that they wished there was a way that they could identify themselves as qualified evaluators, and 62% agreed that “I would likely pursue the requirements of certification.” In addition, at the government level, there was increasing concern about evaluation quality assurance and interest in the professionalization of evaluation (Gussman, 2005). It is the involvement of the government in the process of professional designations that differentiates the Canadian approach and the American approach in pursuit of professionalization of evaluators (Cousins, Cullen, Malik, & Maicher, 2007).

> Evaluation in Canada, similar to most jurisdictions around the globe, is heavily driven by government (Aucoin, 2005; Cousins & Aubry, 2006; Segsworth, 2005). The majority of members of CES, whose roles currently number somewhere in the neighborhood of 1800, are organizationally located in government, mostly at the federal level, but with significant representation from provincial and municipal counterparts…[I]t stands in marked contrast to that of the American Evaluation Association (AEA) most notably with respect to representation

> from the academic sector. We understand that slightly more than 40% of AEA's membership is organizationally located in colleges and universities. (p. 1)

With the involvement of the government, it was more likely for the Canadian Evaluation Society to develop a certification, or even licensure, system for evaluators with less concern about legal implications, which have become one of the biggest roadblocks for the American Evaluation Association (Altschuld, 2007).

The Canadian Evaluation Society National Council initiated the professional designations project in 2006. They issued a request for proposals (RFP) entitled *Fact Finding Regarding Evaluator Credentialing*. "The purpose of this RFP is to attract proposals to produce an action plan that would aid the Canadian Evaluation Society (CES) in establishing a professional credentialing system including a member registry" (Canadian Evaluation Society, 2006, p. 1). Headed by Gerald Halpern, a consortium was formed to deliver a series of papers, including an action plan for professional standards for evaluators (Halpern, Gauthier, & McDavid, 2007), an extensive literature review (Huse & McDavid, 2006), and a report on interview results of 16 professional organizations (Halpern & Long, 2007). In addition, the National Council of the Canadian Evaluation Society also considered major differences in a minority report (Long, 2007).

As response to the consortium's recommendation, the National Council of CES concluded that a system of professional designation should be developed and implemented and the form should be a credentialing system with the possibility to develop an exam-based system of professional certification in the future (Canadian

Evaluation Society National Council, 2007c). The Council agreed to establish a CES Credentialing Board (CES-CB) to oversee the development of criteria and decide the admissibility of individual applicants as Credentialed Evaluators (2007c, p. 10). The Council also allowed a "grand parenting" provision for the transitional period as well as setting up a dispute mechanism that would rule on disputes or appeals to the CES Credentialing Board (2007c, p. 10).

The Professional Designation Core Committee (PDCC) suggests that a credentialed evaluator is a credential holder who "has provided evidence of education and experience required to be a competent evaluator" (Maicher, 2008). Moreover, the PDCC developed a conceptual framework consisting of "three pillars": ethics, professional standards, and evaluator competencies as the underpinning foundation of professional designations (See Figure 2.6) (Buchanan, Maicher, & Kuji-Shikatani, 2008; Cousins, Buchanan, Kuji-Shikatani, & Maicher, 2008).

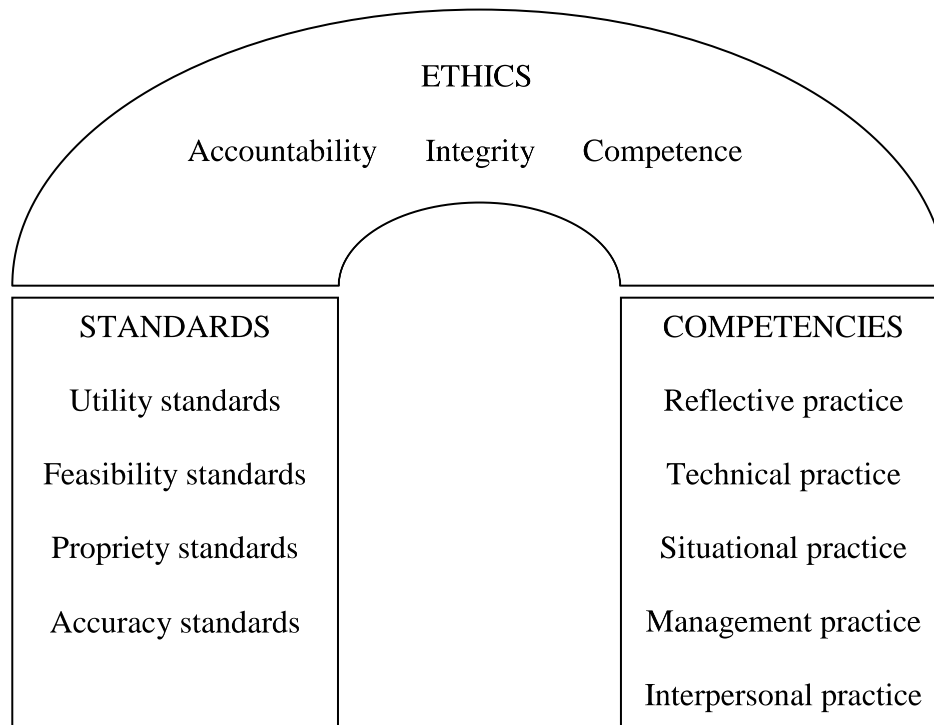

Figure 2.6. Three pillars of professional designations in Canada.

From "Canadian Evaluation Society Professional Designations Project," by H. Buchanan, B. Maicher, & K. Kuji-Shikatani, paper presented at the Annual American Evaluation Association Conference, 2008, Denver, CO.

The CES Guidelines for Ethical Conduct were approved by the National Council in 1996 and reviewed in 2006. Serving as the overarching code of conduct for program evaluators in Canada, the Guidelines are one of the three pillars of professional designation. In March 2008, the CES National Council voted unanimously to adopt the Joint Committee Standards for Program Evaluation as standards for effective Canadian evaluation practice. The standards become the second pillar of professional designation. In May 2009, the CES National Council approved the Competencies for Canadian Evaluation Practice, which serve as the third pillar of professional designation. The

Competencies for Canadian Evaluation Practice build on Stevahn, King, Ghere, and Minnema's (2005) research of *Essential Competencies for Program Evaluators(ECPE)* (Canadian Evaluation Society National Council, 2007c, p. 7), along with a crosswalk of different sources of competencies related to evaluation.

Having been developed in 2001 and revised in 2005, the Essential Competencies for Program Evaluators (ECPE) have been widely used. The authors of the ECPE have long envisioned a comprehensive validation study to determine the validity of the proposed evaluator competencies across the entire field (King, Stevahn, Ghere, & Minnema, 2001; Stevahn, King, Ghere, & Minnema, 2005). The professional designation of the CES also shows the need for a validation study of the ECPE.

Summary

Competence is an abstract construct that refers to broad capacities. It builds on knowledge, skills, and attitudes. Competency, with its plural form, competencies, refers to particular abilities, which can be acquired through academic education and professional practice. The concept of competence was first fully developed in businesses and management with the social efficiency movement. It was used in work and employee selection. Soon after, the concept of competence was introduced to the field of education, and competency-based education became one of the most prevailing approaches to curriculum design. The impact of competency-based education and training subsequently influenced various professions in the United States, such as the health related professions, legal professions, accounting, and technical professions. As an emerging profession, the

field of program evaluation, of course, has been fiercely discussing the possibilities of developing evaluator competencies.

There are three frameworks to understand the concept of competence. They are the differential-psychology approach focusing on human differences, the educational-and-behavioral psychology approach emphasizing developing people from “unconscious incompetence” to “conscious competence,” and the management-science approach underlining job analysis. Empirical research on competencies flourished in management science. Innovative techniques in job analysis were developed to identify job competencies. The classic study design of competencies was developed by Spencer and Spencer (1993) after an extensive review of previous studies.

In the field of evaluation, King, Stevahn, Ghere, and Minnema (2001) independently developed the taxonomy of essential competencies for program evaluators. In 2009, the Canadian Evaluation Society (CES) advanced and adopted Competencies for Canadian Evaluation Practice, which builds on King et al.’s Essential Competencies for Program Evaluators. The Competencies for Canadian Evaluation Practice are used as one of the three pillars to construct professional designation in Canada. Therefore, the Essential Competencies for Program Evaluators are of importance in the professionalization of evaluators.

Having compared the development of the Essential Competencies for Program Evaluators to Spencer and Spencer’s model of a classic study design of competencies, a gap was found that the Essential Competencies for Program Evaluators have been used without validation. It is also the authors’ intention that the Essential Competencies for

Program Evaluators be validated. The next chapter details the methodology of this initial validation study of Essential Competencies for Program Evaluators.

CHAPTER THREE

METHODS

This chapter details this study's methods . As stated in Chapter One, the purpose of this thesis was to investigate the validity of the Essential Competencies for Program Evaluators (ECPE). Accordingly, the research questions are:

1. To what extent are the Essential Competencies for Program Evaluators valid across the field of evaluation?
2. What additions, deletions, or other changes would make the Essential Competencies for Program Evaluators more representative across the field of evaluation?

Based on these questions, this study is nonexperimental since it does not seek to determine cause and effect between independent and dependent variables. It does seek to make systematic observations to collect data (Orcher, 2005; Patten, 2005), and therefore is empirical research to begin the validation process for the Essential Competencies for Program Evaluators. The beginning of this validation process starts with the development of a survey that will be used nation-wide to collect data on the validity of the ECPE. A high quality survey with validity and reliability is one important component to justify the generalization of survey results from a sample to a larger population (Patten, 2005). This section starts with a discussion on the concept of validity, and then reviews the procedures of the survey development, including the plan for data collection and analysis.

Validity

Validity is "an integrated evaluative judgment of the degree to which empirical evidence and theoretical rationales support the *adequacy* and *appropriateness* of *inferences* and *actions* based on test scores or other modes of assessment" (Messick, 1989, p. 13). The term *test score*, according to Messick (1989), not only refers to the numerical score from tests as ordinarily conceived, but also refers to patterns from any means of documenting or observing consistent behaviors or attributes. Messick's definition of validity is helpful in identifying the substance of this study. This study intended to evaluate the degree to which empirical evidence supports the adequacy and appropriateness of inferences from the proposed sixty-one Essential Competencies for Program Evaluators.

Unlike traditional views of validity, Messick (1995b, p. 741) proposes a unified concept of validity that "addresses both score meaning and social values in test interpretation and test use." With the unitary concept of validity, the validation process is based on two important premises. First, to validate is to evaluate both the existing evidence for and the potential consequences of score interpretation and use. Therefore, "what is to be validated is not the test or observation device as such but the inferences derived from test scores or other indicators" (Messick, 1989, p. 13). Second, to validate is to evaluate the degree of fit, not all or none. That is to say, "validity is less a question of 'yes or no,' but instead 'more or less'" (Toal, 2007, p. 47).

Traditionally, there are three categories of validity evidence identified (American Educational Research Association, American Psychological Association, and National

Council on Measurement in Education, 1999), one of which has two sub-categories. They are content-related validity evidence, criterion-related validity evidence, including predictive validity and concurrent validity, and construct-related validity evidence.

Content validity is “an assessment of whether a test contains appropriate content and requires that appropriate processes be applied to that content” (Thorndike, 2005, p. 147). It is deductive and established by showing that the test items are “a sample of a universe in which the investigator is interested” (Cronbach & Meehl, 1955, p. 282). Messick (1989) described content validity as providing support for “the domain relevance and representativeness of the test instrument,” but saw it as playing a limited role in validation because it does not provide direct evidence “in support of inferences to be made from test scores” (p. 17).

Criterion validity is the degree to which test scores correlate with some chosen criterion measures of job or academic success. “The higher the correlation, the more effective the test is as a predictor and the higher is its criterion-related validity” (Thorndike, 2005, p. 157). If the test is given before the criterion is obtained, the validity is called *predictive validity*; if the test score and criterion score are determined at the same time, the validity is called *concurrent validity* (Cronbach & Meehl, 1955, p. 282). Criterion validity is of use in many applied contexts, for example, when an employer uses an accepted measure of job performance as a criterion for hiring (Kane, 2001). However, a problem may arise regarding the validity of the criterion: “the need for a well-defined and demonstrably valid criterion measure” (Kane, 2001, p. 320).

Considered as an alternative to criterion and content validity, construct validity was introduced in 1950s by Meehl and Challman in response to the need of the American Psychological Association Committee on Psychological Tests to broaden the definition of validity (Kane, 2001). Cronbach and Meehl state that "construct validation is involved whenever a test is to be interpreted as a measure of some attribute or quality which is not 'operationally defined'" (1955, p. 282). According to Cronbach and Meehl (1955),

> Construct validity is not to be identified solely by particular investigative procedures, but by the orientation of the investigator…Construct validity must be investigated whenever no criterion or universe of content is accepted as entirely adequate to define the quality to be measured. Determining what psychological constructs account for test performance is desirable for almost any test. (p. 282)

Cronbach and Meehl (1955) identified three methodological principles in the development of construct validity: the need for extended analysis, the need for an explicit statement of the proposed interpretation, and the need to consider alternate interpretations (Kane, 2001, p. 324). For criterion validity, test scores are simply compared to the criterion scores. For content validity, the characteristics of the measurement procedure are evaluated based upon expert opinions. For construct validity, validation calls for an extended analysis including the development of a theory, measurement procedures, and hypothesis testing (Kane, 2001). Therefore, the validation of construct validity entails the development of a proposed interpretation or several competing interpretations of test scores, which was not a concern for either criterion or content validity (Kane, 2001).

Rather than treating construct validity as one type of validity, Loevinger (1957) went a step further, suggesting construct validity as a general approach to include all kinds of validity evidence. She articulated the comprehensive nature of construct validity by arguing that "since predictive, concurrent, and content validities are all essentially ad hoc, construct validity is the whole of validity from a scientific point of view" (Loevinger, 1957, p. 636). The view of construct validity embracing all forms of validity evidence soon became popular in the 1960s and 1970s, and Messick adopted it as a general framework for his unified validity (Messick, 1989).

Kane (2001, pp. 328-329) summarized four features of Messick's unitary view of validity. First, validity involves an evaluation of the overall plausibility of a proposed interpretation or use of test scores. It is the interpretation including inferences and decisions that is validated, not the test or the test scores. Second, the proposed interpretation will involve an extended analysis of inferences and assumptions and will involve both a rationale for the proposed interpretation and a consideration of possible competing interpretations. Third, validation includes the evaluation of the consequences of test uses. The 1999 *Standards for Educational and Psychological Testing* state that those who propose to use a test score in a particular way, e.g., to make a particular kind of decision, are expected to justify the use, and proposed use is generally justified by showing that the positive consequences outweigh the anticipated negative consequences (American Educational Research Association, American Psychological Association, and National Council on Measurement in Education). Fourth, validity is an integrated, or unified, evaluation of the interpretation.

"The essence of unified validity is that the appropriateness, meaningfulness, and usefulness of score-based inferences are inseparable and that the integrating power derives from empirically grounded score interpretation" (Messick, 1995b, p. 747), therefore, meaning and values are integral parts of the concept of validity. To address both value implications and social consequences, the unified validity framework is of two interconnected facets. "One facet is the source of justification of the testing, being based on appraisal of either evidence or consequences. The other facet is the function or outcome of the testing, being either interpretation or use" (Messick, 1989, p. 20). The facet for justification is crossed with the facet for function or outcome, resulting in a four-fold classification highlighting both meaning and values in both test interpretation and test use, as shown in Figure 3.1.

Function/Outcome Justification	Test Interpretation	Test Use
Evidential Basis	Construct Validity (CV)	CV + Relevance/Utility (R/U)
Consequential Basis	CV + Value Implications	CV + R/U + Social Consequences

Figure 3.1. Facets of validity.

Modified by the author from "Validity," by S. Messick. In R. L. Linn (Ed.), *Educational Measurement* (3rd ed.). Macmillan Publishing Company, 1989.

…[T]he evidential basis of test interpretation is construct validity. The evidential basis of test use is also construct validity, but is buttressed by evidence for the relevance of the test to the specific applied purpose and for the utility of the test in the applied setting. The consequential basis of

test interpretation is the appraisal of the value implications of the construct label, of the theory underlying test interpretation, and of the ideologies in which the theory is embedded… Finally, the consequential basis of test use is the appraisal of both potential and actual social consequences of the applied testing. (Messick, 1989, p. 20)

Using the unified theory of validity implies that a validation study will present multiple forms of evidence to demonstrate the extent to which a test is adequate and appropriate for specified uses. Based on Messick (1989, 1995a, 1995b), the six criteria of validity, briefly summarized below, should accumulate evidence of the validity of the meaning and use of test scores:

1. Content-related evidence supports the extent to which all aspects of the construct are covered by the items on the test.
2. Substantive-related evidence supports the extent to which the items reflect an individual's actual performance.
3. Structural-related evidence supports the extent to which scoring models should be consistent with theoretical and rational structures underlying the construct.
4. Generalizability-related evidence supports the extent to which the items reflect the same construct across settings or groups.
5. Externally-related evidence supports the extent to which the construct is related or not related to other variables.
6. Consequence-related evidence supports the extent to which there is evidence of positive and negative consequences.

In Messick's view, these six criteria of validity should be addressed simultaneously in the validity argument. They cannot be dealt with in isolation, which makes this unified validity theory different from other traditional validity theories. As a result, any validation study needs to collect evidence to meet these six criteria of validity. Table 3.1 shows what kind of evidence is needed to address Messick's six criteria of validity with regard to the Essential Competencies for Program Evaluators (ECPE).

Table 3.1

Evidence Needed to Validate the Essential Competencies for Program Evaluators Using Messick's Six Criteria

Messick's Criteria of Validity	**Questions to Be Asked about the ECPE**
Content-related Evidence	To what extent do the ECPE measure an evaluator's competence?
Substantive-related Evidence	To what extent are the ECPE inclusive of all necessary competencies for an evaluator to conduct evaluation studies?
Structural-related Evidence	To what extent does the scoring model[4] of the ECPE match the theoretical and rational structures of evaluator competencies?
Generalizability-related Evidence	To what extent are the ECPE applicable to evaluators who practice in various content areas?
Externally-related Evidence	To what extent does evaluators' competence correlate with other measures of evaluator competencies other than the ECPE?
Consequence-related Evidence	To what extent would the use or the interpretations of the ECPE not have negative consequences for evaluators?

[4] The scoring model refers to the Essential Competencies for Program Evaluators Self-Assessment Tool (Stevahn, L., King, J., Ghere, G., & Minnema, J., 2004).

To answer the six above-mentioned questions regarding the validity of the Essential Competencies for Program Evaluators (ECPE), a large scale comprehensive validation study is called for, which is beyond the scope of this thesis. The current study took the first step toward a full-fledged validation study of the ECPE. It developed instruments to initially validate the ECPE to ground a nation-wide validation study in the future.

Procedures

The two main methods used to conduct this study were survey and interview. To collect evidence from a range of respondents, including practicing evaluators, faculty members in evaluation studies, and program management staff, an online survey was developed. “A survey is a method of gathering information from a number of individuals, known as a sample, in order to learn something about the larger population from which the sample is drawn” ("What is a web survey?" 2003, para. 1). This web-based survey was hosted by the College of Education and Human Development at the University of Minnesota for the duration of the study. This survey collected respondents’ opinions and attitudes toward the ECPE as well as their demographic information. Interviews were conducted after the implementation and the analysis of the online survey. Purposeful sampling was used to select participants for the interviews. All of the interviewees had worked in the field of evaluation for an extensive period of time, close to 15 years on average. They were considered “experts” in their respective content areas. Therefore, the interviews were also called “expert interviews” in this study.

Survey items and the interview protocol were developed with Messick's six criteria of validity in mind to collect evidence needed to validate the Essential Competencies for Program Evaluators (ECPE). Table 3.2 shows the research methods used to answer questions about the validity of the ECPE.

Table 3.2

Research Methods Used to Answer Questions about the Validity of the Essential Competencies for Program Evaluators

Messick's Criteria of Validity	Questions to Be Asked about the ECPE	Method
Content-related Evidence	To what extent do the ECPE measure an evaluator's competence?	Survey
Substantive-related Evidence	To what extent are the ECPE inclusive of all necessary competencies for an evaluator to conduct evaluation studies?	Survey
Structural-related Evidence	To what extent does the scoring model of the ECPE match the theoretical and rational structures of evaluator competencies?	N/A[5]
Generalizability-related Evidence	To what extent are the ECPE applicable to evaluators who practice in various content areas?	Interview
Externally-related Evidence	To what extent does evaluators' competence correlate with other measures of evaluator competencies other than the ECPE?	Interview
Consequence-related Evidence	To what extent would the use or the interpretations of the ECPE not have negative consequences for evaluators?	Interview

Survey Development Process

A robust process generated the survey. A seven-point scale was originally created and then reduced to a five-point scale after an extensive think aloud process.

[5] The Essential Competencies for Program Evaluators Self-Assessment Tool was not examined as part of this study.

A seven-point scale. A seven-point scale was developed to solicit respondents' opinions on the appropriateness of each proposed competency. The respondents would have the opportunity to look at each of the proposed competencies and indicate the extent to which they agreed the competencies were necessary for program evaluators according to the seven-point scale as follows:

1 = Not at all necessary

2 = Very slightly necessary

3 = Slightly necessary

4 = Moderately necessary

5 = Strongly necessary

6 = Very strongly necessary

7 = Extremely necessary

This scale started with "Not at all necessary" and went up six steps to "Extremely necessary." A scale like this one starting with the absence of certain attributes of a construct and increasing one way toward all of the attributes is considered a unipolar scale as opposed to a bipolar scale, which has a middle point with steps that increase in both ways as they move toward the opposite ends (Pope, 2005; Henning, 2009; Krosnick & Tahk, 2011). The survey in question used a unipolar scale because it provides the respondent with a greater range of subtlety in terms of response options (Church & Waclawski, 2001, p. 75).

The web-based survey was originally designed to use the seven-point scale because the advantage of a seven-point scale over a four- or five-point scale is its ability

to detect smaller differences (Veldhuyzen Van Zanten et al., 2006). In addition, the seven-point scale seems to have the most gain in reliability among all other scales. Nunnally (as cited in Lewis, 2001) commented on the number of scale steps. He believed "a greater number of scale steps will enhance scale reliability but with rapidly diminishing returns. As the number of scale steps increases from two to twenty, there is an initially rapid increase in reliability that tends to level off at about seven steps" (p. 149).

In addition to this seven-point scale, the survey was designed to ask respondents to comment in an open-ended question format. Respondents could share their concerns and suggestions, if any, toward any particular competencies. Demographic information was collected at the end of the survey to inform the characteristics of the sample of evaluators.

To reduce measurement error in the development of the survey, the configuration and language of this web-based survey were closely examined. According to Dillman (2007), measurement error occurs when inaccurate or un-interpretable answers are obtained as a result of poor question wording or questions being presented in a way that is vague and confusing. He also identified three other common types of errors in survey research: coverage error, sampling error, and non-response error, which will be discussed in the section on survey implementation.

To ensure a clear online format, the survey was divided to be shown on fourteen webpages to minimize the need to scroll up and down on the computer screen. The seven-point scale was designed to appear on top of a group of competencies in each webpage of

the survey so that respondents could easily refer to the scale as they rated each competency. To ensure the quality of the questions and further reduce measurement error, a think aloud process was used.

Think aloud. Think aloud is also known as concurrent cognitive interview (Dillman, 2007; Willis, 2004). It is used to pilot instruments so that changes can be made to improve them. A think aloud is used to tell how questions are working. Therefore, the think aloud focuses on the questions, not on the participant's actual responses. Participants are asked to verbalize their thinking process so that their understanding of the questions is "seen." The think aloud reflects a participant's reactions to the survey's content, format, and ease of understanding.

Taking into consideration that there are 61 competencies on the list of the Essential Competencies for Program Evaluators (ECPE), a think aloud with all 61 competencies would have been too long for an interview. To keep the think aloud within appropriate length, the competencies were divided into four groups with 16 items in the first group and 15 each in the other three groups respectively. The groups of items were formed according to their sequential order in the ECPE. As a result, each participant was able to evaluate at least two out of six categories of the ECPE. Two participants were invited to look at the same group of items. Therefore, 8 people in total were recruited for the think aloud process. These 8 people were graduate students who had studied either in evaluation studies or measurement research. They were also practicing evaluators who had experience with program evaluation. They were familiar with evaluator competencies.

Table 3.3 shows the grouping of items for the think aloud process. It took about 45 minutes for each person to evaluate 15 or 16 items.

Table 3.3

Grouping of Items for Think Aloud

Group	Items Numbers	Number of Items
1	1.1-2.10	16
2	2.11-3.5	15
3	3.6-4.8	15
4	4.9-6.6	15
	Total	61

The process of think alouds was recorded upon consent. The recordings were transcribed and analyzed. Four themes surfaced from these 8 think aloud interviews.

First, participants tended to draw on their personal experience when answering the survey questions, whereas the intention of the survey was to consider "evaluators in general." All participants related to their personal experience with evaluation when they reflected on the proposed competencies. Although some respondents did notice the wording of "whether or not they do so regularly" in the directions, it appeared that they still thought about their own experience when answering the survey questions. In addition, it was more likely for them to rate low on the scale if the competencies did not apply to them. It was also likely for them to rate low on the scale if they did not understand what particular competencies meant. As a result, two sentences were added into the scale statement. They were "Don't think about your own experience" and "Think about evaluators in general." These two sentences were bold, italicized, and put in bright yellow

to draw respondents' attention. Because they were added into the scale, the two sentences reappeared with the scale on every webpage of the online survey.

Second, participants tended to critique the wording of the competencies without knowing that the wording of each competency was exactly as it appeared in the ECPE. One participant suggested "putting in the instruction that the wording here reflects the exact wording of the competencies, because I want to know whether you made up this wording or whether it really reflects the actual competencies that were published. Are you trying to test her [King et al.'s] competencies, or you are trying to make up new competencies?" As a result, the phrase, "the wording of competencies was exactly taken from the ECPE," was added to the general directions that preceded the scale.

Third, demographic questions were the most confusing to the 8 respondents. In addition, some of the choices were not mutually exclusive. For example, when asked about the job title, one participant said, "I am an evaluator; I am a graduate student; I am also program staff." She hesitated about which option to choose. When asked about the most time spent relevant to job function, one participant was frustrated, "I do research; I do data collection; I do data analysis; and I do program evaluation… But really all other pieces are also part of program evaluation." As a result, the demographic questions underwent major changes. Some questions were taken out, some questions were completely re-written, and some questions were given different options.

Last, seven out of eight participants suggested changing the seven-point scale to a five-point scale. "I was struggling with the seven-point scale. Can I really so finely answer those questions? What's the difference between 'very slightly necessary' and

‘slightly necessary’? I am not sure what that means.” “Seven points are a lot of points. I am not sure if I know the distance between ‘very slightly’ and ‘slightly’”. “…[I]t seems to be quite a bit of mental work to make a distinction between ‘very slightly’ and ‘slightly,’ or ‘very strongly’ and ‘strongly.’ What’s the distance between ‘very strongly’ and ‘extremely’?” As a result, a five-point scale was finally developed.

A five-point scale. Empirically, the five-point scale was suggested not only from the think aloud data, but also from Krosnick and Tahk’s (2011) research. They found that bipolar scales performed best with seven points, whereas unipolar scales performed best with five. The scale in question was a unipolar scale. Therefore, the seven-point scale, which was originally designed, was replaced by the five-point scale as follows:

1 = Not at all necessary

2 = Slightly necessary

3 = Moderately necessary

4 = Strongly necessary

5 = Extremely necessary

Compared to the original survey, the revised survey was improved in such areas as the number of steps in the scale, the web-based presentation, clearer language, and revamped demographic questions. The revised survey was sent to the same 8 think aloud participants for further comments. Two of them wrote short emails back indicating that they did not have new comments. Six of them did not provide further information. [See the entire survey in Appendix A]

IRB Process

This study was submitted to the Independent Review Board (IRB) for approval before any implementation of the survey and other data collection. The IRB application was accepted with exemption from full IRB review (Study Number: 0906E68141). A protocol amendment was then filed to reflect the revision of the instruments. The IRB approved this change in protocol request.

Survey Implementation

To conduct the study, the revised survey was administered to a sample group of evaluators in Minnesota. According to Dillman (2007, p. 11), there are four types of errors common in survey research: measurement error, coverage error, sampling error, and nonresponse error. The measurement error was addressed in the development of the survey. The other three common errors were addressed in the implementation of the survey.

Coverage error. Coverage error is a result of not allowing all members of the survey to have an equal or known chance of being sampled for participation in the survey. To eliminate the coverage error, the survey was sent to the entire listserv of the Minnesota Evaluation Association (MN EA). The MN EA is the local affiliate of the American Evaluation Association. There are more than 200 contacts on the listserv, which include evaluators from various fields of study (e.g., public health, education, policy analysis, research methods, and etc.) who work in various types of organizations (e.g., in academia, government, independent consultant, non-profit organization, etc.). In addition, to facilitate the expert interviews later on, a group of 12 “experts” was identified

to receive the survey. Some of the experts were on the MN EA listserv, but only received the survey once to avoid duplication. As a result, the total number of evaluators in the sample was 208.

Sampling error. Sampling error is a result of surveying only some, not all elements of the survey population. The sampling error was minimized in this study because the survey was sent to the entire MN EA listserv except for those without a valid email. The survey was successfully sent to 208 people, including 196 on the MN EA listserv and 12 experts.

Nonresponse error. Nonresponse error is the result of people who respond to a survey being different from those who did not respond in ways relevant to the study. To reduce nonresponse error, a letter of invitation was sent along with the link to the survey. The letter of invitation explained the intention for the survey, the implication to the field of evaluation, and the call for participation. The letter of invitation also offered alternative formats of the survey to make sure everyone had a chance to participate. The timing for sending out the survey was carefully chosen to avoid long holidays and events (i.e., tax return due date, beginning or ending of a school year). Up to three reminders were sent to those who had not responded. As a result, 93 out of 196 people on the listserv and 9 out of 12 experts responded to the survey. The survey response rate was 49%. As a rule of thumb, the acceptable response rate for e-mail and online surveys is 30-40% (Hamilton, 2009; The University of Texas at Austin, 2011). This study had a higher response rate than the average.

Interview Process

Nine interviews were conducted after the implementation and the analysis of the online survey. As explained earlier, purposeful sampling was used to select participants for the interviews. All of the interviewees had worked in the field of evaluation for an extensive period of time, close to 15 years on average. They were considered experts in their respective content areas: public health, education, and non-profits. There were three experts in each of these three areas.

The reason for interviewing expert evaluators in three different content areas was to look at their unique perspectives. In other words, would they respond differently to the interview questions due to their practicing in different content areas? The interview protocol included the background information of the interviewee, the use of evaluator competencies in their particular fields, and the positive and negative consequences if the Essential Competencies for Program Evaluators were to be institutionalized. [See the complete interview protocol in Appendix B]

The interviews were semi-structured. Although there were predetermined questions regarding interviewees' evaluation practice in relation to evaluator competencies, new questions were brought up to probe or clarify an answer. All of the interviews were recorded and then transcribed to inform the eventual validity process of the Essential Competencies for Program Evaluators.

Summary

Validity is “an integrated evaluative judgment of the degree to which empirical evidence and theoretical rationales support the *adequacy* and *appropriateness* of *inferences* and *actions* based on test scores or other modes of assessment” (Messick, 1989, p. 13). This study was designed to evaluate the degree to which empirical evidence supports the adequacy and appropriateness of the inferences of the proposed sixty-one Essential Competencies for Program Evaluators. Using the unified theory of validity implies that a validation study will present multiple forms of evidence to demonstrate the extent to which a test is adequate and appropriate for specified uses. Based on Messick (1989, 1995a, 1995b), six criteria of validity should accumulate evidence of the validity of the meaning and use of test scores. The current study takes the first step toward a full-fledged validation study of the ECPE. It developed instruments to preliminarily validate the ECPE to shed light on the nation-wide validation study in the future.

CHAPTER FOUR

RESULTS

Chapter Four presents the survey and interview results, along with a brief explanation of how the results are relevant to the research questions. The extent to which the results provide evidence of validity will be discussed in Chapter 5.

Survey Results

The survey was designed to gather information on the extent to which respondents believed that program evaluators need to be able to demonstrate essential competencies for program evaluators (ECPE), whether or not they do so regularly. The survey asked respondents to rate each competency on the list of the ECPE as extremely necessary, strongly necessary, moderately necessary, slightly necessary, or not at all necessary.

Respondents

The survey collected respondents' experience with program evaluation and their background information. Among the 102 respondents, 89 resided in Minnesota, and 76 were female. Seventy percent of the respondents directly worked in the field of evaluation as consultants, directors, or staff. Ten percent of them were program personnel, including project managers, program directors and staff. The rest of the respondents were research associates, policy and planning analysts, and a couple of retirees. Almost all of them worked for non-business organizations; for example, non-profits, government, and colleges/universities were the top three types of organizations that two-thirds of them

worked for. Only 7 people indicated that they had worked for businesses or industry. Other organizations included K-12 school districts, research firms, public health organizations, and foundations.

Less than a quarter of the respondents (23) were self-employed, but 45 of them indicated that they had more than 15 years of experience in the field of evaluation. Table 4.1 shows how the respondents labeled themselves regarding their experience with program evaluation.

Table 4.1

Respondents' Self-reported Labeling of Their Experience

Respondents	Count
Mastery	23
Skilled	43
Proficient	26
Novice	9
No answer	1
Total	102

As far as education is concerned, the respondents were highly trained. There were 42 (41%) respondents with doctoral degrees. Seventeen of them held their doctoral degree in evaluation studies. There were 49 respondents who had earned master's degrees in various fields. Almost all of them had received training in at least three or more areas such as evaluation theory, evaluation practice, research methods/design, quantitative methods, qualitative methods, mixed methods, and measurement. Fourteen respondents were graduate students. In addition, five respondents had earned a certificate in program evaluation. Most of them (67%) associated themselves with the American Evaluation

Association and/or its local affiliate Minnesota Evaluation Association[6] among other professional associations.

Frequency Counts

The survey was sent to a sample of 208 evaluators in Minnesota through the MN EA listserv including 12 people who were identified as expert evaluators. One hundred and two of them, including 9 expert evaluators, responded to the survey. Therefore, the response rate was 49%. The results of the survey are detailed in a frequency table, shown as Table 4.2. For example, for the ECPE 1.1, "Applies professional evaluation standards," 51 respondents or 50.00% of all respondents rated it as extremely necessary; 42 respondents or 41.18% rated it as strongly necessary; 7 respondents or 6.86% rated it as moderately necessary; 2 respondents or 1.96% rated it as slightly necessary; and no respondents rated it as not at all necessary.

[6] Not all people who are on the Minnesota Evaluation Association (MN EA) listserv are members of MN EA.

Table 4.2 Table of the Frequency Results

Competency	Extremely necessary (5)		Strongly necessary (4)		Moderately necessary (3)		Slightly necessary (2)		Not at all necessary (1)		Weighted Mean	Weighted SD
	Counts	%	Counts	%	Counts	%	Counts	%	Counts	%		
1.1	51	50.00	42	41.18	7	6.86	2	1.96	0	0.00	4.39	0.81
1.2	91	89.22	10	9.80	0	0.00	1	0.98	0	0.00	4.87	0.51
1.3	28	27.45	43	42.16	27	26.47	4	3.92	0	0.00	3.93	0.96
1.4	81	79.41	19	18.63	1	0.98	1	0.98	0	0.00	4.76	0.59
1.5	40	39.22	44	43.14	13	12.75	5	4.90	0	0.00	4.17	0.96
1.6	9	8.82	21	20.59	47	46.08	20	19.61	5	4.90	3.09	1.09
2.1	50	49.02	35	34.31	16	15.69	1	0.98	0	0.00	4.31	0.89
2.2	32	31.37	55	53.92	14	13.73	1	0.98	1	0.98	4.17	0.84
2.3	28	27.45	60	58.82	12	11.76	1	0.98	1	0.98	4.11	0.80
2.4	35	34.31	49	48.04	15	14.71	2	1.96	1	0.98	4.13	0.89
2.5	17	16.67	40	39.22	34	33.33	10	9.80	1	0.98	3.61	1.02
2.6	15	14.71	44	43.14	29	28.43	10	9.80	4	3.92	3.55	1.10
2.7	58	56.86	35	34.31	8	7.84	1	0.98	0	0.00	4.47	0.79
2.8	56	54.90	41	40.20	4	3.92	1	0.98	0	0.00	4.49	0.72
2.9	58	56.86	39	38.24	4	3.92	1	0.98	0	0.00	4.51	0.72
2.10	45	44.12	38	37.25	13	12.75	4	3.92	2	1.96	4.18	1.04
2.11	49	48.04	44	43.14	5	4.90	4	3.92	0	0.00	4.35	0.87
2.12	48	47.06	44	43.14	5	4.90	5	4.90	0	0.00	4.32	0.90
2.13	55	53.92	38	37.25	6	5.88	3	2.94	0	0.00	4.42	0.85
2.14	68	66.67	31	30.39	2	1.96	1	0.98	0	0.00	4.63	0.67

Competency	Extremely necessary (5)		Strongly necessary (4)		Moderately necessary (3)		Slightly necessary (2)		Not at all necessary (1)		Weighted Mean	Weighted SD
	Counts	%	Counts	%	Counts	%	Counts	%	Counts	%		
2.15	34	33.33	36	35.29	24	23.53	7	6.86	1	0.98	3.93	1.08
2.16	24	23.53	48	47.06	24	23.53	4	3.92	2	1.96	3.86	0.99
2.17	39	38.24	49	48.04	13	12.75	1	0.98	0	0.00	4.24	0.81
2.18	62	60.78	33	32.35	5	4.90	2	1.96	0	0.00	4.52	0.79
2.19	59	57.84	38	37.25	4	3.92	1	0.98	0	0.00	4.52	0.72
2.20	3	2.94	19	18.63	39	38.24	26	25.49	15	14.71	2.70	1.15
3.1	37	36.27	47	46.08	14	13.73	3	2.94	1	0.98	4.14	0.93
3.2	39	38.24	41	40.20	18	17.65	3	2.94	1	0.98	4.12	0.97
3.3	50	49.02	36	35.29	12	11.76	4	3.92	0	0.00	4.29	0.95
3.4	62	60.78	32	31.37	6	5.88	2	1.96	0	0.00	4.51	0.80
3.5	33	32.35	41	40.20	19	18.63	9	8.82	0	0.00	3.96	1.07
3.6	36	35.29	46	45.10	13	12.75	6	5.88	1	0.98	4.08	1.00
3.7	27	26.47	31	30.39	36	35.29	7	6.86	1	0.98	3.75	1.07
3.8	38	37.25	47	46.08	13	12.75	4	3.92	0	0.00	4.17	0.92
3.9	16	15.69	43	42.16	33	32.35	8	7.84	2	1.96	3.62	1.02
3.10	52	50.98	38	37.25	10	9.80	2	1.96	0	0.00	4.37	0.85
3.11	57	55.88	37	36.27	6	5.88	2	1.96	0	0.00	4.46	0.80
3.12	53	51.96	38	37.25	9	8.82	2	1.96	0	0.00	4.39	0.84
4.1	11	10.78	30	29.41	39	38.24	12	11.76	10	9.80	3.20	1.22
4.2	37	36.27	46	45.10	13	12.75	6	5.88	0	0.00	4.12	0.97
4.3	26	25.49	38	37.25	25	24.51	10	9.80	3	2.94	3.73	1.16
4.4	65	63.73	29	28.43	5	4.90	3	2.94	0	0.00	4.53	0.84

Competency	Extremely necessary (5)		Strongly necessary (4)		Moderately necessary (3)		Slightly necessary (2)		Not at all necessary (1)		Weighted Mean	Weighted SD
	Counts	%	Counts	%	Counts	%	Counts	%	Counts	%		
4.5	40	39.22	41	40.20	16	15.69	5	4.90	0	0.00	4.14	0.98
4.6	35	34.31	45	44.12	15	14.71	7	6.86	0	0.00	4.06	1.01
4.7	56	54.90	36	35.29	9	8.82	1	0.98	0	0.00	4.44	0.80
4.8	33	32.35	45	44.12	20	19.61	4	3.92	0	0.00	4.05	0.95
4.9	29	28.43	33	32.35	27	26.47	10	9.80	3	2.94	3.74	1.19
4.10	29	28.43	40	39.22	22	21.57	9	8.82	2	1.96	3.83	1.12
4.11	32	31.37	40	39.22	20	19.61	8	7.84	2	1.96	3.90	1.11
4.12	49	48.04	45	44.12	7	6.86	1	0.98	0	0.00	4.39	0.76
5.1	53	51.96	32	31.37	15	14.71	2	1.96	0	0.00	4.33	0.92
5.2	41	40.20	42	41.18	13	12.75	6	5.88	0	0.00	4.16	0.99
5.3	38	37.25	42	41.18	14	13.73	7	6.86	1	0.98	4.07	1.04
5.4	28	27.45	45	44.12	20	19.61	5	4.90	4	3.92	3.86	1.12
5.5	25	24.51	45	44.12	24	23.53	8	7.84	0	0.00	3.85	1.01
6.1	61	59.80	37	36.27	3	2.94	1	0.98	0	0.00	4.55	0.70
6.2	72	70.59	27	26.47	2	1.96	1	0.98	0	0.00	4.67	0.65
6.3	35	34.31	48	47.06	13	12.75	4	3.92	2	1.96	4.08	1.00
6.4	31	30.39	44	43.14	19	18.63	5	4.90	3	2.94	3.93	1.09
6.5	55	53.92	34	33.33	9	8.82	3	2.94	1	0.98	4.360	0.94
6.6	48	47.06	40	39.22	10	9.80	4	3.92	0	0.00	4.29	0.92

Means and Standard Deviations

Although the Likert-type scale like the one used in this survey does not possess a normal probability distribution because of its categorical nature, in practice it is common to compute means of the scale since they can reveal the direction of the average response. The standard deviation (SD) is also important as it indicates the average distance from the mean. A low standard deviation indicates that most responses cluster around the mean; a high standard deviation indicates more variation in the responses.

Furthermore, the scale in question is a unipolar scale (i.e., extremely necessary is coded as 5; strongly necessary is coded as 4; moderately necessary is coded as 3; and so on). The responses to such a unipolar scale are often treated as ordinal data. Therefore, weighted means are reported instead of regular means as well as weighted standard deviations[7]. Table 4.3 details the weighted means and weighted standard deviations for each competency in the ECPE. The competencies are organized according to descending weighted means.

Table 4.3

Weighted Means and Weighted Standard Deviations

No.	Label	Competency	Weighted Mean	Weighted SD
1	1.2	Acts ethically and strives for integrity and honesty in conducting evaluations	4.87	0.51
2	1.4	Respects clients, respondents, program	4.76	0.59

[7] The weighted standard deviation is calculated by using the following formula:

$$\mathrm{SDw} = \sqrt{\sum_{i=1}^{N} Wi\,(Xi - \bar{X}w)^2 / \frac{(N'-1)\sum_{i=1}^{N} Wi}{N'}}$$

where w_i is the weight for the ith observation, N' is the number of non-zero weights, and $\bar{X}w$ is the weighted mean of the observations.

No.	Label	Competency	Weighted Mean	Weighted SD
		participants, and other stakeholders		
3	6.2	Uses verbal/listening communication skills	4.67	0.65
4	2.14	Interprets data	4.63	0.67
5	6.1	Uses written communication skills	4.55	0.70
6	4.4	Communicates with clients throughout the evaluation process	4.53	0.84
7	2.18	Reports evaluation procedures and results	4.52	0.79
8	2.19	Notes strengths and limitations of the evaluation	4.52	0.72
9	3.4	Serves the information needs of intended users	4.51	0.80
10	2.9	Identifies data sources	4.51	0.72
11	2.8	Develops evaluation design	4.49	0.72
12	2.7	Frames evaluation questions	4.47	0.79
13	3.11	Remains open to input from others	4.46	0.80
14	4.7	Identifies needed resources for evaluation, such as information, expertise, personnel, instruments	4.44	0.80
15	2.13	Analyzes data	4.42	0.85
16	3.12	Modifies the study as needed	4.39	0.84
17	1.1	Applies professional evaluation standards	4.39	0.81
18	4.12	Presents work in a timely manner	4.39	0.76
19	3.10	Respects the uniqueness of the evaluation site and client	4.37	0.85
20	6.5	Facilitates constructive interpersonal interaction (teamwork, group facilitation, processing)	4.36	0.94
21	2.11	Assesses validity of data	4.35	0.87
22	5.1	Aware of self as an evaluator (knowledge, skills, dispositions)	4.33	0.92
23	2.12	Assesses reliability of data	4.32	0.90
24	2.1	Understands the knowledge base of evaluation (terms, concepts, theories, assumptions)	4.31	0.89
25	3.3	Identifies the interests of relevant stakeholders	4.29	0.95
26	6.6	Demonstrates cross-cultural competence	4.29	0.92
27	2.17	Provides rationales for decisions throughout the evaluation	4.24	0.81
28	2.10	Collects data	4.18	1.04
29	1.5	Considers the general and public welfare in evaluation practice	4.17	0.96
30	3.8	Attends to issues of evaluation use	4.17	0.92
31	2.2	Knowledgeable about quantitative methods	4.17	0.84

No.	Label	Competency	Weighted Mean	Weighted SD
32	5.2	Reflects on personal evaluation practice (competencies and areas for growth)	4.16	0.99
33	4.5	Budgets an evaluation	4.14	0.98
34	3.1	Describes the program	4.14	0.93
35	2.4	Knowledgeable about mixed methods	4.13	0.89
36	3.2	Determines program evaluability	4.12	0.97
37	4.2	Negotiates with clients before the evaluation begins	4.12	0.97
38	2.3	Knowledgeable about qualitative methods	4.11	0.80
39	3.6	Examines the organizational context of the evaluation	4.08	1.00
40	6.3	Uses negotiation skills	4.08	1.00
41	5.3	Pursues professional development in evaluation	4.07	1.04
42	4.6	Justifies cost given information needs	4.06	1.01
43	4.8	Uses appropriate technology	4.05	0.95
44	3.5	Addresses conflicts	3.96	1.07
45	2.15	Makes judgments	3.93	1.08
46	6.4	Uses conflict resolution skills	3.93	1.09
47	1.3	Conveys personal evaluation approaches and skills to potential clients	3.93	0.96
48	4.11	Conducts the evaluation in a nondisruptive manner	3.90	1.11
49	5.4	Pursues professional development in relevant content areas	3.86	1.12
50	2.16	Develops recommendations	3.86	0.99
51	5.5	Builds professional relationships to enhance evaluation practice	3.85	1.01
52	4.10	Trains others involved in conducting the evaluation	3.83	1.12
53	3.7	Analyzes the political considerations relevant to the evaluation	3.75	1.07
54	4.9	Supervises others involved in conducting the evaluation	3.74	1.19
55	4.3	Writes formal agreements	3.73	1.16
56	3.9	Attends to issues of organizational change	3.62	1.02
57	2.5	Conducts literature reviews	3.61	1.02
58	2.6	Specifies program theory	3.55	1.10
59	4.1	Responds to requests for proposals	3.20	1.22

No.	Label	Competency	Weighted Mean	Weighted SD
60	1.6	Contributes to the knowledge base of evaluation	3.09	1.09
61	2.20	Conducts meta-evaluations	2.70	1.15

As the data indicate, there are forty-three competencies that had weighted means greater than 4 points. They are considered “Strongly necessary” to “Extremely necessary.” The first ten competencies that had weighted means greater than 4.50 could actually be rounded to 5.00 points. They are considered “Extremely necessary.” Weighted means are rounded to the nearest whole number prior to categorization.

There are fifteen competencies that had weighted means between 3.50 and 4.00 points. If rounded to the whole number, they can be considered at 4 points. Therefore, in a general sense, these fifteen competencies can be considered “Strongly necessary.” As a result, respondents rated 58 out of the 61 competencies “Strongly necessary” or more.

The remaining three competencies are “Responds to requests for proposals” with a weighted mean 3.20; “Contributes to the knowledge base of evaluation,” 3.09; and “Conducts meta-evaluations,” 2.70. If rounded to the whole number, they can be considered at 3 points. Therefore, they were viewed as “Moderately necessary.” None of the competencies were, on average, rated as “not at all necessary” or “slightly necessary.”

Interview Results

As explained earlier, nine participants were purposefully selected to be interviewed. Interviewees were selected based upon the assumption that they would respond differently to the interview questions due to their practice in different content

areas, i.e., public health, education, and non-profits. All of the interviewees had worked in the field of evaluation for an extensive period of time, close to 15 years on average. In each interview, the interviewees talked about their background information, their use of evaluator competencies in their particular fields, and the positive and negative consequences if the Essential Competencies for Program Evaluators (ECPE) were to be institutionalized.

To hone the coding skills of the interview, extensive practice was used to ensure the accuracy and consistency throughout the coding process. The researcher and another doctoral student in evaluation independently coded one selected interview. The coding was then compared and consensus was reached. This exercise helped to calibrate the coding of the rest of the interviews. The results of the interviews are detailed below.

Interviewees

For two of the groups, the interview results showed that the distinction across the content areas of their practice did not hold. Instead, the evaluators talked about their experience from their different roles in conducting evaluation-related activities. Therefore, six of the interviewees were re-grouped according to these job functions. Education remained a distinct category. But the other two categories (public health, non-profits) were restructured into “formal organization” for participants who worked in one and “independent consultant” for participants who worked in this role. There were still three participants identified in each of these new groups.

Comments on the ECPE

Each interviewee responded to questions about his or her use of evaluator competencies in a particular field (education) or from the perspective of evaluators in a formal organization or as independent consultants.

Across the nine interviews, professional practice was considered "primary" as identified by one interviewee. "That's one of those that you just kind of build into what you do," another interviewee commented. More comments on competencies in the category of professional practice focused on evaluation standards, ethics, and personal approaches. For example, "…the underpinning for me is the professional practice, my professionalism. …We adhere to the evaluation standards, and then considering ethical issues, integrity, making sure people understand our approach…," and "I adhere to the standards, the evaluation standards, the guiding principles, and list those out for people."

Interpersonal competence, as expected, turned out to be very important for evaluators as voiced by all interviewees. One interviewee referred to interpersonal competence as "soft skills," which can be very helpful not only in conducting evaluations, but also in "building relationship with clients," "communicating with stakeholders," and "keeping a pulse on the community."

All of the interviewees agreed that systematic inquiry encompassed essential functions of evaluation. They also believed that no one can be proficient in all competencies listed. As a result, they were reportedly more likely to build a team of evaluators who had complementary skills to each other. It was more so for evaluators as independent consultants. They were more likely to pull in a collaborator or two for larger

projects. As for evaluators in formal organizations, it was less an issue since the organization had hired people based upon complementary skill sets. As far as school districts were concerned, people who worked in the department of research, assessment, and evaluation mainly performed their duties using quantitative methods with testing and assessment data unless there was funding for enrichment programs that needed to be evaluated. Almost all interviewees reportedly did not do meta-evaluation "because of time and money constraints."

The response to situational analysis varied by fields and roles of evaluators. Evaluators in school districts seldom reported conducting situational analysis simply because they are already "situated" in their districts. One interviewee explained,

> I wouldn't be the one who would need to describe the program. I mean, I would maybe need people to describe their program to me a little bit so I could help them figure out what they could potentially ask or evaluate and what kind of data might be available. A lot of times people come up to me and ask me what data [I have] available already.

On the contrary, independent evaluators reported that they conduct situational analysis quite often. "I think it's really important to get enmeshed with what's going on and what they're doing and what they're trying to do, what people think of it. To learn a lot of the context before I start," one interviewee recalled. Another interviewee said, "…[M]y belief is we've got to build that capacity in the organization."

Depending upon whether the role of an evaluator was internal or external, situational analysis was sometime used. For example, one interviewee, who is an internal evaluator for an organization, reported that "… this is not done as much," while an evaluator from a research and evaluation firm working as an external evaluator of a project would act more like an independent consultant. As a matter of fact, the ability to conduct situational analysis was considered as a higher level of task that experienced evaluators would process. "[A]s I'm thinking about the differential level of experience and the trajectory of inexperienced evaluators becoming more experienced, it's probably a capacity to do some of the situational analysis…" In addition, "… doing the situational analysis, the responsibility may fall to the senior person."

By contrast, evaluators who work as independent consultants reported conducting project management all the time. One interviewee commented on several competencies listed under project management. "[If] you're gonna be out on your own, you'd better have a good budget process." "Negotiating before we begin, before we sign the contract. I think it's really important for them to understand what they're getting…" "So it's really important to make sure they understand that, always have a formal agreement, always!" For evaluators who worked in a firm or an organization, the project management often fell on the senior staff. "That's a primary role that I have [as a senior research associate]," another interviewee said.

All interviewees thought highly about reflective practice, but acknowledged they could do better. Resources, mainly time and money, reportedly became the crucial factor if they were to pursue reflective practice. Evaluators in professional organizations

reported doing a better job at regular staff development and training. This was reflected by one interviewee from a formal organization.

> [W]e took advantage of formal staff development opportunities, then we had a monthly reflection meeting that was intended to focus on improving our skill organizationally, but also individual practice. We had mentoring relationships established internally, so that every evaluator had at least one or more senior evaluator to turn to.

Evaluators who worked in different fields or served in different roles did demonstrate some difference in emphases in the six areas of the ECPE as shown in Table 4.4. Evaluators as independent consultants reportedly demonstrated all six areas of the ECPE, with the least emphasis on systematic inquiry, which they usually achieved through collaboration with other independent evaluators. Evaluators in school districts were more likely to work on research and assessment in addition to evaluation projects. They tended to focus on systematic inquiry using quantitative methods. Evaluators in formal organizations reported often working in teams that created the capacity to execute all six areas of competencies.

Table 4.4

Summary of Expert Interview Results

ECPE Category	Evaluators in Education	Evaluators in Formal Organizations	Evaluators as Independent Consultants
Professional practice	Primary	Primary	Primary
Systematic inquiry	Essential Heavy reliance on testing and assessment data	Essential Team building	Essential Collaboration with one another
Situational analysis	Seldom Internal evaluators in school district	Sometimes Depending upon the role of evaluator as internal or external	Frequently Evaluation capacity building
Project management	Not so much unless managing tests	Team leader or senior staff's responsibility	All the time
Reflective practice	Wish to be able to do more regularly Heavily dependent upon district budget	Could do better Regular staff meetings and rotate attending conferences	Trying to keep up Informally and locally Dependent upon funding
Interpersonal competence	Very important	Very important	Very important

Certification/Credentialing

Except for one person who supported the idea of having some kind of certification/credentialing system in place, the rest of interviewees had mixed feelings about this idea. The positive and negative consequences of institutionalizing the ECPE they discussed in forms of certification/credentialing are detailed in Table 4.5.

Table 4.5

Pros and Cons of Certification/Credentialing

Pros	Cons	Considerations
• Merit in the maturity of the profession • This would standardize a set of minimal skills that someone could expect from an evaluator • Client would be more confident to hire evaluators	• Might keep some people out of the profession who have great skills and potential • Application and continuing education would cost money • Some who have degrees in evaluation are very skilled, but lack program background • Might be a way to show off to people as a way to get another job	• Broad entry • Depends on the set up • Start with people who are just entering the field

On the advantage side, interviewees reported, "…the field is moving toward [a] credential." They believed that "it brings more credibility to the field" and "bring[s] more standardization to the field to have a highly qualified evaluator." All interviewees agreed that certification and/or credentialing extended benefit to the clients. "I think that's in the best interest of client[s]," one interviewee said. Another said, "If you're somebody who's trying to hire people, knowing that at least at some point in time they were able to prove their competencies would be a helpful thing." Most of them believed, "there's probably more positive about the concept than there would be negative."

On the disadvantage side, most interviewees had problems with exclusivity, the cost, the potential for over-specialization, and viability. For example, one interviewee said, "Some of the best evaluators that I've worked with have been program people first

and have learned program evaluation as a result of their program development or their program implementation skills." Therefore, "…establishing a credential [system] would keep some people out of the profession who have great skills and the potential to be very good evaluators."

As far as the cost is concerned, one interviewee said, "If you had to pay $500 just to go through the board process, and then they require you to do two conferences within three years to keep your credential, that's a huge expense." She continued to give an example of how the cost would influence a credentialed and a non-credentialed evaluator.

> So if I were to not choose to do this [go through the credentialing process], and I am $20 an hour cheaper than anybody else, but they have the credential, would they get upset at me because I am not credentialed and I am undercutting their business because I am cheaper? And I explain to the client that I've gotten a master's in it [evaluation]. I could go through the credential process, but that would cost me $2,000 a year and I am passing the savings on to you by being $20 cheaper than they are. If you want the credential, then go ahead and take them or you can come back to me. So would it cause that sort of a rift between evaluators?

One interviewee worried about over-specialization. She thought instead of "… [becoming] a distinct discipline where evaluation is done in relationship to school programs, public health programs, re-development programs, correctional situation," certification would "… just isolate itself [the field of evaluation] too much." "I just think that's too rigid," she added. Another interviewee considered that "the diversity of practice

might be limited." "You'd have to fit into certain molds to be an evaluator. Sometimes people who come in from outside the field are really great. They come at it a different way, but they do add value."

Speaking of viability, one interviewee expressed her concerns about the credibility of this designation.

> If you put in a system like this, you may have people say, "I am now a credentialed evaluator." But they've just barely made it through and became credentialed and that was ten or fifteen years ago and maybe they haven't done anything with it since then. So I think for some people, being able to say "a credentialed evaluator" will be more of something they can just show off to people as a way of being able to get another job.

She continued to say, "I've been doing this kind of work for 18 years. In my case I just finished my Ph.D. in evaluation in education. Really? Do I need to go through another hoop to be credentialed? It would seem like, really? More?"

The concerns about the credentialing system resulted in a great deal of consideration of how to set up the system as one interviewee said, "… It all depends on how it gets set up." Another interviewee cautioned that "there are just lots of pieces to think through if you're going to put something that big in place." It seemed like there were more questions to be answered. For example, "It becomes like a checklist?" "Who gets to judge it?" "How do I get certified?" "Are you going to come around with me and watch me do the situational analysis?"

Findings

Comparing and contrasting the survey results with the interview results, the following findings were observed. In the survey results, the first ten competencies all had weighted means greater than 4.50. They could actually round to 5.00 points and be considered "Extremely necessary." Table 4.6 lists the first ten competences.

Table 4.6

First Ten Competences

No.	Label	Competency	Weighted Mean	Weighted SD
1	1.2	Acts ethically and strives for integrity and honesty in conducting evaluations	4.87	0.51
2	1.4	Respects clients, respondents, program participants, and other stakeholders	4.76	0.59
3	6.2	Uses verbal/listening communication skills	4.67	0.65
4	2.14	Interprets data	4.63	0.67
5	6.1	Uses written communication skills	4.55	0.70
6	4.4	Communicates with clients throughout the evaluation process	4.53	0.84
7	2.18	Reports evaluation procedures and results	4.52	0.79
8	2.19	Notes strengths and limitations of the evaluation	4.52	0.72
9	3.4	Serves the information needs of intended users	4.51	0.80
10	2.9	Identifies data sources	4.51	0.72

The first two competencies, "Acts ethically and strives for integrity and honesty in conducting evaluations" and "Respects clients, respondents, program participants, and other stakeholders" were about professional practice. They were mirrored in the expert interviews as "primary."

Interpersonal competence was considered very important by all interviewees, which was also reflected in the survey results. "Uses verbal/listening communication

skills" and "Uses written communication skills" were placed at #3 and #5 respectively in the first ten competencies.

There were four competencies under systematic inquiry that were included in the first ten competencies. Likewise, all interviewees agreed that systematic inquiry encompassed essential functions of evaluation. However, they also pointed out that nobody can be proficient in all competencies in systematic inquiry. As a result, they believed building a team of evaluators who had complementary skills to each other was the best way to go.

The interview results showed that evaluators reported various levels of proficiency in situational analysis and project management competencies. Depending upon whether an evaluator was internal or external, at entry level or senior level, situational analysis was sometime used; depending upon the field and the role of an evaluator, project management competencies were desired by some. However, the survey results showed that any competency about relationship and interaction with clients and intended users was rated very important. "Communicates with clients throughout the evaluation process" and "Serves the information needs of intended users" placed at #6 and #9 respectively served as the evidence from the survey results.

These first ten competencies came from five categories of the ECPE, all except for reflective practice. A close look at competencies under reflective practice showed that as a whole, they did not rank high. Table 4.7 shows the placement of reflective practice competencies. The interview results shed lights on the reason why reflective practice did not rank high. Although all interviewees thought highly about reflective practice, they

acknowledged that resources, mainly time and money, prevented them from reflecting on their practice. They all hoped to do better if possible.

Table 4.7

Placement of Reflective Practice Competencies

No.	Label	Competency	Weighted Mean	Weighted SD
22	5.1	Aware of self as an evaluator (knowledge, skills, dispositions)	4.33	0.92
32	5.2	Reflects on personal evaluation practice (competencies and areas for growth)	4.16	0.99
41	5.3	Pursues professional development in evaluation	4.07	1.04
49	5.4	Pursues professional development in relevant content areas	3.86	1.12
51	5.5	Builds professional relationships to enhance evaluation practice	3.85	1.01

The least necessary competency among all 61 competencies, rated by survey participants, was "conducts meta-evaluation," which had a weighted mean of 2.70. Almost all interviewees reportedly did not do meta-evaluation either. The reported reason for not conducting meta-evaluation was "because of time and money constraints." According to the majority of survey and interview respondents, meta-evaluation as an evaluation concept and practice is seldom well-understood, and is likely misunderstood by many. A few experts in the interviews did see value in conducting meta-evaluation, but expressed that it was not part of their job responsibilities.

Summary

Chapter Four presented the survey and interview results. The survey was designed to gather information on the extent to which respondents believed that program evaluators need to be able to demonstrate essential competencies for program evaluators (ECPE), whether or not they do so regularly. The survey was sent to a sample of 208 evaluators in Minnesota through the MN EA listserv. One hundred and two of them (49%) responded to the survey.

Weighted means and weighted standard deviations were reported. Fifty-eight out of the 61 competencies were considered "Strongly necessary" or more. The remaining three competencies were viewed as "Moderately necessary." None of the competencies were, on average, rated as "not at all necessary" or "slightly necessary."

The interview results revealed that professional practice and interpersonal competence were considered critical by all. Although evaluators were reportedly rarely proficient in all systematic inquiry competencies, they reportedly were most likely to collaborate on these essential functions of evaluation. Evaluators who worked in different fields or served in different roles demonstrated variations in conducting situational analysis and project management. All interviewees thought highly about reflective practice, but acknowledged that they could do better in this area. Almost all interviewees reported they did not do meta-evaluation.

Regarding certification and/or credentialing, almost all interviewees had mixed feelings. All interviewees agreed that certification and/or credentialing would potentially extend benefits to their clients. Some interviewees expressed their reservations with its

exclusivity, the cost, the potential for over-specialization, and viability. The concerns about a potential certification and/or credentialing system in the U.S. resulted in a great deal of consideration of possible system configurations.

Chapter Five discusses the evidence of validity of the ECPE as a result of the survey and interview results.

CHAPTER FIVE

DISCUSSION AND CONCLUSION

Evidence of Validity

As stated in Chapter One, the purpose of this thesis was to develop instruments and preliminarily validate the Essential Competencies for Program Evaluators (ECPE). Chapter Five uses evidence of validity from the survey and interview results to answer the following research questions:

1. To what extent are the Essential Competencies for Program Evaluators valid across the field of evaluation?
2. What additions, deletions, or other changes would make the Essential Competencies for Program Evaluators more representative across the field of evaluation?

As stated earlier, validity is "an integrated evaluative judgment of the degree to which empirical evidence and theoretical rationales support the *adequacy* and *appropriateness* of *inferences* and *actions* based on test scores or other modes of assessment" (Messick, 1989, p. 13). The term *test score*, according to Messick (1989), not only refers to the numerical score from tests as ordinarily conceived, but also to patterns from any means of documenting or observing consistent behaviors or attributes. Messick's definition of validity was helpful in interpreting the survey and interview results of this study. These survey and interview results were used to evaluate the degree

to which they supported the adequacy and appropriateness of the inferences of the proposed sixty-one essential competencies for program evaluators.

Using the unified theory of validity advanced by Messick (1989, 1995a, 1995b), multiple forms of evidence were collected to demonstrate the extent to which the ECPE was adequate and appropriate for its intended use. Messick (1989, 1995a, 1995b) suggested there were six criteria of validity that should be addressed simultaneously in a validity argument. The survey and interview results provided evidence of validity in five of these six criteria. Structural-related evidence, which supports the extent to which scoring models should be consistent with theoretical and rational structures underlying the construct, was not examined in this study. Although the Essential Competencies for Program Evaluators have been turned into a self-assessment tool (Ghere, King, Stevahn, & Minnema, 2006) with which an evaluator could self-evaluate each of his or her competencies on a scale of zero to six, the extent to which the scoring model of the ECPE matches the theoretical and rational structures of evaluator competencies was not part of this dissertation.

To determine the degree of validity, the strength of each of these five criteria of evidence was measured using a simple rating of "strong, limited, or mixed." Similar to the process Toal (2007) used, if the survey and interview results indicated that the ECPE adequately provided evidence addressing the specific criteria of validity, then the evidence was considered "strong"; if the survey and interview results provided some evidence, but not adequately, then the evidence was considered "limited"; if the survey and interview results showed inconsistent evidence, then the evidence was considered

“mixed.” These ratings were used to form the evaluative judgment of whether the ECPE is valid across the field of evaluation. Table 5.1 shows the strength of evidence for each criterion of the evidence.

Table 5.1

Summary of the Evidence

Messick’s Criteria of Validity	**Questions to Be Asked about the ECPE**	**Method**	**Strength of Evidence**
Content-related Evidence	To what extent do the ECPE measure an evaluator’s competence?	Survey	Strong
Substantive-related Evidence	To what extent are the ECPE inclusive of all necessary competencies for an evaluator to conduct evaluation studies?	Survey	Strong
Generalizability-related Evidence	To what extent are the ECPE applicable to evaluators who practice in various content areas?	Interview	Mixed
Externally-related Evidence	To what extent does evaluators’ competence correlate with other measures of evaluator competencies other than the ECPE?	Interview	Limited
Consequence-related Evidence	To what extent would the interpretations of the ECPE not have negative consequences for evaluators?	Interview	Strong

Content-related Evidence

Content-related evidence supports the extent to which the items reflect an individual’s actual performance, in this case, the “ideal” performance of evaluators in general. As mentioned earlier, the survey was designed to gather information on the extent to which respondents *believed* that program evaluators need to be able to

demonstrate essential competencies whether or not they do so regularly. In the context of the ECPE, content-related evidence was collected of the extent to which the ECPE reportedly measured an evaluator's competence.

Compared to the 58 competencies that were considered "Strongly necessary" or "Extremely necessary," three competencies—"Responds to requests for proposals," "Contributes to the knowledge base of evaluation," and "Conducts meta-evaluations"—ranked lower. If rounded to a whole number, they could still be viewed as "Moderately necessary." None of the competencies were, on average, rated as "not at all necessary" or "slightly necessary." Therefore, there is strong evidence that the ECPE reportedly measured an evaluator's "ideal" competence.

However, there is one more consideration to this claim. Some respondents found it a challenge "to answer questions on behalf of all evaluators" rather than their own experience. The survey sought to specify the "ideal" list of essential competences for all evaluators "whether or not they do so regularly." It is likely that some respondents reported what they do as "necessary" and what they do the least as "moderately necessary" or "slightly necessary." "Responds to requests for proposals," "Contributes to the knowledge base of evaluation," and "Conducts meta-evaluations" were reportedly less-frequently conducted activities by an evaluator. As a result, it is very likely that the instruments had measurement error. Considering that correcting the measurement error would result in increased rigor of the evidence, not the other way around, the content-related evidence was still considered strong despite the potential measurement error.

Substantive-related Evidence

Substantive-related evidence supports the extent to which all aspects of the construct are covered by the items on the instrument, in this case, the ECPE. In the context of the ECPE, the substantive-related evidence was collected on the extent to which the ECPE were inclusive of all necessary competencies for an evaluator to conduct evaluation studies. The survey gathered information on the extent to which respondents believed that program evaluators needed to be able to demonstrate 61 essential competencies, whether or not they did so regularly. Fifty-eight out of the 61 competencies (95%) were considered "Strongly necessary" or "Extremely necessary." The remaining three competencies were viewed as "Moderately necessary." None of the competencies were, on average, rated as "not at all necessary" or "slightly necessary." Figure 5.1 shows the weighted mean distribution of 61 competencies. The mean of these weighted means was 4.14, which is "Strongly necessary." The small weighted standard deviation (0.40) indicated that measurement error was minimal, and most weighted means clustered around 4.14.

In addition, both the survey and the interview asked respondents if they had any comments regarding any addition, deletion, or other changes to the essential competencies. The results did not yield significant additions or deletions to the ECPE. Therefore, the substantive-related evidence was considered strong.

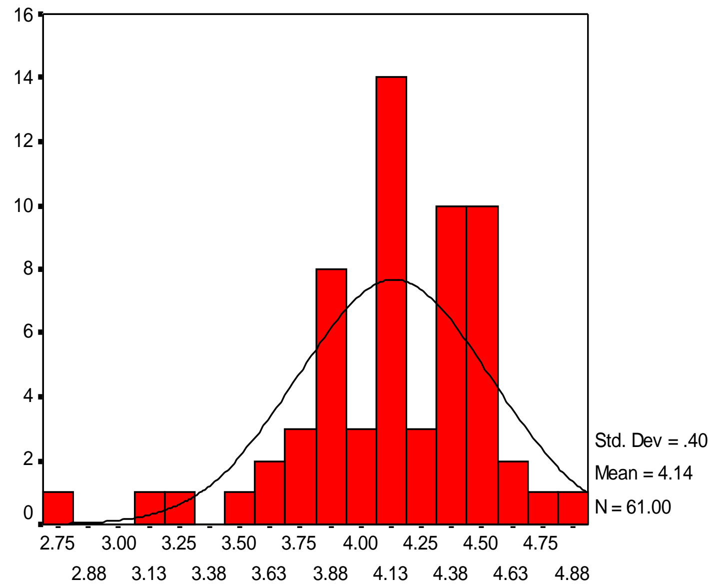

Figure 5.1. Weighted mean distribution of 61 competencies.

Generalizability-related Evidence

Generalizability-related evidence supports the extent to which the items reflect the same construct across settings or groups. In the context of the ECPE, the generalizability-related evidence was collected through expert interviews. Expert interviewees were originally selected according to their practicing content areas, i.e., public health, education, and non-profits. The intent was to see whether they would respond the same or differently to the interview questions, hence, whether or not the ECPE is generalizable across settings or groups.

In reality, the interview data necessitated re-grouping of these 9 interviewees. Six of the interviewees were re-grouped into "formal organization" for participants who worked in one and "independent consultant" for participants who worked in this role. Education remained a distinct group. Although these groups were not the ones that were

intended in the design phase of the study, they did represent evaluators who worked in different settings. Therefore, the interview data provided evidence for generalization.

As reported earlier, evaluators who worked in different fields or served in different roles did demonstrate different emphases in the six areas of the ECPE. Evaluators as independent consultants reportedly demonstrated all six areas of the ECPE with the least emphasis on systematic inquiry, which they usually achieved through collaboration with other independent evaluators. Evaluators in formal organizations reported often working in teams that created the capacity to execute all six areas of competencies while individuals do not have the capacity to fulfill all 61 competencies. Evaluators in school districts were more likely to focus their work on research and assessment with some effort in program evaluation. They did not think of situational analysis and project management as their essential functions.

As one can see, there is a continuum of how the ECPE is generalizable across groups. On one end stands the individual evaluator, who cannot fulfill all competencies; on the other end stands the assembly of evaluators, as a group, who want to build a team with complementary skills that cover all essential competencies. Therefore, the generalizability-related evidence was mixed. This suggests that ideally, evaluators, as individuals or a group, need to have all competencies to perform all kinds of evaluation tasks. However, in practice, individual evaluators or evaluators in different roles and functions often choose to specialize in a few evaluation tasks.

Externally-related Evidence

Externally-related evidence supports the extent to which the construct is related or not related to other variables. In the context of the ECPE, the externally-related evidence was collected on the extent to which evaluators' competence correlated with other measures of evaluator competencies rather than the ECPE. In this regard, the survey and the interview did ask respondents if they could think of any other competencies that needed to be added to the ECPE. Neither the survey nor the interview results yielded significant additional information to the ECPE, but one interviewee mentioned that the ECPE needed to be updated to reflect the latest development in system theory. Therefore, there is some externally-related evidence, but it may not be adequate due to no other reasons sought. So, the externally-related evidence was limited.

Consequence-related Evidence

Consequence-related evidence supports the extent to which there is evidence of positive and negative consequences. In the context of the ECPE, the consequence-related evidence was collected through interview data about the use or the interpretations of ECPE having a potential negative impact on evaluators. Most respondents thought of the ECPE as "a great tool to reflect." One interviewee said, "This is a very complete list. They [the creators of the ECPE] did a very good job of going through and thinking about the things you need to do in the different areas." One student's comment shed light on how the ECPE could be used to develop training programs to meet the need of upcoming evaluators: "As a student, it was actually really helpful to go through the competencies in

this way." Some business-driven evaluators thought of the ECPE having the potential to be a "marketing tool." The consequence-related evidence was strong and positive.

One step further, the interviewer also asked respondents both the positive and negative consequences of institutionalizing the ECPE in forms of certification/ credentialing. Regarding certification and/or credentialing, almost all interviewees had mixed feelings. All interviewees agreed that certification and/or credentialing extended benefit to the clients. Some interviewees expressed their reservations with potential exclusivity, the cost, the potential for over-specialization, and viability. The concerns about the potential certification and/or credentialing system in the U.S. resulted in a great deal of consideration of a system configuration.

Using Messick's (1989, 1995a, 1995b) unified theory of validity, evidence was collected against five criteria of validity. Content-related evidence, substantive-related evidence, and consequence-related evidence were strong. The extent that the ECPE measures an evaluator's competence was strong; the ECPE was strongly inclusive of all necessary competencies for one to conduct evaluation studies; and there was strong evidence that the use or the interpretations of the ECPE did not have negative consequences for evaluators. The generalizability-related evidence was mixed. Often individual evaluators or evaluators in different roles and functions choose to specialize in a few evaluation tasks, which require some but not all competencies; while other evaluators need to be proficient in all competencies to perform all required tasks. The externally-related evidence was limited due to inadequate responses. Table 5.2

summarizes the strength of conclusion for each question in relation to its validity of evidence.

Table 5.2

Summary of the Conclusion

Messick's Criteria of Validity	Questions to Be Asked about the ECPE	Strength of Evidence	Strength of Conclusion
Content-related Evidence	To what extent do the ECPE measure an evaluator's competence?	Strong	Strong
Substantive-related Evidence	To what extent are the ECPE inclusive of all necessary competencies for an evaluator to conduct evaluation studies?	Strong	Strong
Generalizability-related Evidence	To what extent are the ECPE applicable to evaluators who practice in various content areas?	Mixed	Mixed
Externally-related Evidence	To what extent does evaluators' competence correlate with other measures of evaluator competencies other than the ECPE?	Limited	Inconclusive
Consequence-related Evidence	To what extent would the interpretations of the ECPE not have negative consequences for evaluators?	Strong	Strong and positive

Implications for Future Research

Structural-related evidence, which supports the extent to which scoring models should be consistent with theoretical and rational structures underlying the construct, was not examined in this study, but it could be a natural next step for further validating the ECPE as a self-assessment Tool (Ghere, King, Stevahn, & Minnema, 2006). The self-assessment tool enables an evaluator to self-evaluate each of his or her competencies on a scale of zero to six. A future study could examine the extent to which the scoring model of the ECPE matches the theoretical and rational structures of evaluator competencies.

In this study, interviewees were selected based upon the assumption that they would respond differently to the interview questions due to their practice in different content areas, i.e., public health, education, and non-profits. For two of the groups, the interview results showed that this distinction did not hold. Instead, the evaluators talked about their experience from their different roles in conducting evaluation-related activities. Therefore, six of the interviewees were re-grouped according to these job functions. Education remained a distinct category. But the other two categories (public health, non-profits) were restructured into “formal organization” for participants who worked in one and “independent consultant” for participants who worked in this role. Further studies could examine program evaluators’ functions and roles.

The survey was designed to gather information on the extent to which respondents believed that program evaluators need to be able to demonstrate essential competencies whether or not they do so regularly. However, there was a challenge for some respondents to answer questions on behalf of all evaluators rather than their own experience. Although reminded, some respondents continued to report what they do as “necessary” and what they do the least as “Moderately necessary” or “Slightly necessary.” Further studies could conduct an analysis on how much personal experience could influence one’s judgment as an evaluator.

This study is an initial study to preliminarily validate the Essential Competencies for Program Evaluators. A full-fledged nation-wide validation study could expand upon and further review the process and results of this initial study. For example, the sample

size could be enlarged from people who are on the Minnesota Evaluation Association mailing list to the American Evaluation Association membership.

Implications for Practice

The validation of the Essential Competencies for Program Evaluators has many implications for practice that are important for students, new evaluators, experienced evaluators, individual evaluators, organizations that hire evaluators, and academic and training programs. Such validation is also crucial to the professionalization of the practice.

Students who are still learning the basics of evaluation will benefit greatly from knowing the full range competencies that will be helpful to their practice and will be potentially required by future employers. The validated ECPE will also alert them to the knowledge, skills, and attitudes that may be required but not covered adequately by their coursework. They may be of a non-academic nature, for example, interpersonal competence, but crucial to the effective practice of evaluation nonetheless. The validated ECPE will also help them to think about areas that they might choose to specialize in, based both on their interest and the relative importance of the competency, as indicated in the validation.

New evaluators who have been trained in program evaluation will be able to use the ECPE to help them determine the competencies in which they are most interested as well as the competencies in which they are most qualified, thus enabling them to make more informed choices when seeking employment within organizations, working with their employer to determine the type of work that they would be best suited for, or

deciding to venture out on their own in a consultative setting. The validated ECPE may also assist them in assessing the strengths and weakness of the organizations they are with, enabling them to both improve their own work and to make a contribution to the organization as a whole by working with others to ensure that all of the competencies are covered by at least someone within the organization and someone on each project. The validated ECPE could also be helpful to a new evaluator in making effective continuing education choices by selecting additional training in competencies in which they perceive themselves to be least proficient and in competencies most relevant to the work they do.

The reactions of the interviewees in this validation study show that even professionals with fifteen or more years of experience recognize the importance of having validated ECPE to use in their practice. Since program evaluation is still a young field, experienced evaluators who are practicing today have taken a variety of paths to get to their current positions. Some have more of an academic background while others were predominantly trained on-the-job. Some work on their own and need to be "generalists," while others work within an organization and have evolved into a specialization within the field, while relying on colleagues for proficiency in certain competencies not necessary to their specialty. All may benefit from exposure to and knowledge of the validated ECPE. For example, the generalist solo practitioner may be aware that he or she needs to "do everything," but would undoubtedly benefit from having a more specific knowledge of what "everything" entails, in order to either improve on weaknesses or bring other evaluators into a project when additional proficiency in certain competencies is called for. A specialist within an organization—even one who does not believe it is

necessary to be proficient in all of the competencies due to his or her specialization—still needs to know that all off the competencies are covered on a given project and needs to choose coworkers or outside assistance accordingly.

Organizations that hire evaluators have a definite need for validated ECPE in order to hire evaluators who cover all of the competencies well and to effectively assign employees to projects. Both of these needs are ongoing, as new assignments are given and employee turnover causes gaps in coverage that need to be filled.

Academic and training programs need to ensure that their programs and materials cover all of the competencies and would benefit from noting the extent to which practicing professionals value each of the competencies. These programs will also benefit from reviewing the competencies to evaluate how well they are providing training in both competencies that lend themselves to academic and traditional classroom training and those that are more difficult to impart in that manner. To the extent that some competencies are based more on attitudes and personal skills than on knowledge of information, these programs would do well to make students aware of the importance of those competencies even if "teaching" them is difficult.

Professionalization of the field of evaluation is not possible without an adequate description of knowledge, skills, and attitudes that are necessary to the practice of program evaluation. The validated ECPE can provide the basis for competency-based certification/credentialing. Professional organizations (e.g., American Evaluation Association) could provide academic program accreditation based on a list of validated

evaluator competencies. The validated ECPE would provide exactly the type of information that is required for the professionalization of evaluation.

Conclusion

This study investigated two research questions with regards to the Essential Competencies for Program Evaluators developed by King et al. (2001) and revised by Stevahn at al. (2005):

1. To what extent are the Essential Competencies for Program Evaluators valid across the field of program evaluation?
2. What additions, deletions, or other changes would make the Essential Competencies for Program Evaluators more representative across the field of evaluation?

To validate the ECPE is different from validating test scores in that the ECPE is a list of attributes of competence which are not numerical scores from tests as ordinarily conceived by many validation studies. According to Messick (1989), patterns from any means of documenting or observing consistent behaviors or attributes are treated the same way as test scores in a validity argument. This study provided such an example of conducting a validation study beyond validating test scores.

In this study, the researcher spent a great deal of energy on the development of instruments—the web-based survey, which has undergone a robust think aloud process, as well as the expert interview protocol with purposeful sampling. According to Messick's (1989, 1995a, 1995b) unified theory of validity, evidence was collected using

the survey and interviews and compared to the six criteria of validity: content-related evidence, substantive-related evidence, structural-related evidence, generalizability-related evidence, externally-related evidence, and consequence-related evidence.

The survey and interview results provided evidence of validity for five of these six criteria. The degree of validity of these five criteria was measured using a simple rating of "strong, limited, or mixed." The degree of validity of the ECPE in the content-related evidence, the substantive-related evidence, and the consequence-related evidence was strong. The generalizability-related evidence was mixed for two reasons. First, individual evaluators or organizations that hire a group of evaluators may build teams with complementary competencies to be able to conduct various evaluation tasks across disciplines and content areas; individual evaluators may not need every competency. Second, the results suggested that one can never cover all competencies. People may choose to specialize in one area or a few. Externally-related evidence was limited because only limited evidence was collected.

This initial study provided evidence that speaks to the validity of the Essential Competencies for Program Evaluators. To further improve the ECPE, one could investigate the last-rated three competencies ("Responds to requests for proposals," "Contributes to the knowledge base of evaluation," and "Conducts meta-evaluation"), further collect evidence on the externally-related evidence, and conduct validation at a national level.

References

Allen, C. R. (1919). *The instructor, the man, and the job: A hand book for instructors of industrial and vocational subjects*. Philadelphia: Lippincott.

Altschuld, J. W. (1999a). The case for a voluntary system for credentialing evaluators. *American Journal of Evaluation, 20*(3), 507-517.

Altschuld, J. W. (1999b). The certification of evaluators: Highlights from a report submitted to the Board of Directors of the American Evaluation Association. *American Journal of Evaluation, 20*(3), 481-493.

Altschuld, J. W. (2005). Certification, credentialing, licensure, competencies, and the like: Issues confronting the field of evaluation. *Canadian Journal of Program Evaluation, 20*(2), 157-168.

Altschuld, J. W. (2007). *Critical examination of "An action plan for the Canadian Evaluation Society with respect to professional standards for evaluators"*. Paper presented at the annual meeting of the American Evaluation Association, Baltimore, Maryland.

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *The standards for educational and psychological testing*. Washington, DC: American Psychological Association.

American Evaluation Association. (2004). Guiding principles for evaluators. Retrieved March 24, 2008, from http://www.eval.org/Publications/GuidingPrinciples.asp

American Evaluation Association Task Force on Guiding Principles for Evaluators. (1995). Guiding principles for evaluators. In W. R. Shadish, D. L. Newman, M. A. Scheirer & C. Wye (Eds.), *Guiding principles for evaluators. New directions for program evaluation* (Vol. 62, pp. 19-26). San Francisco: Jossey-Bass.

Anastasi, A., & Urbina, S. (1997). *Psychological testing* (7th ed.). Upper Saddle River, NJ: Prentice Hall.

Bassellier, G., Reich, B. H., & Benbasat, I. (2001). Information technology competence of business managers: A definition and research model. *Journal of Management Information Systems, 17*(4), 159-182.

Becker, H., & Kirkhart, K. (1981). The standards: Implications for professional licensure and Accreditation. *American Journal of Evaluation, 2*, 153-156.

Bickman, L. (1997). Evaluating evaluation: Where do we go from here? *Evaluation Practice, 18*, 1-16.

Bickman, L. (1999). AEA, bold or Timid? *American Journal of Evaluation, 20*(3), 519-520.

Bird, A., & Osland, J. (2004). Global competencies: An introduction. In H. Lane, M. Maznevksi, M. E. Mendenhall & J. McNett (Eds.), *Handbook for Global Managers*. Malden, MA: Blackwell.

Bobbitt, J. F. (1912). The elimination of waste in education. *The Elementary School Teacher, 12*(6), 259-271.

Borys, S., Gauthier, B., Kishchuk, N., & Roy, S. (2005). *Survey of evaluation practice and issue in Canada.* Paper presented at the Joint Canadian Evaluation Society and American Evaluation Association conference, Toronto.

Boyatzis, R. E. (1982). *The competent manager: A model for effective performance*. New York: Wiley.

Buchanan, H., Maicher, B., & Kuji-Shikatani, K. (2008, November 6). *Canadian Evaluation Society professional designations project.* Paper presented at the Annual American Evaluation Association Conference, Denver, CO.

Canadian Evaluation Society. (1999). Essential skills series in evaluation. Retrieved January 15, 2004, from http://www.evaluationcanada.ca

Canadian Evaluation Society. (2004). Essential skills series in evaluation. Retrieved March 27, 2008, from http://www.evaluationcanada.ca/site.cgi?s=3&ss=31&_lang=en

Canadian Evaluation Society. (2006). Request for proposals: Fact finding regarding evaluator credentialing. Retrieved November 26, 2007, from http://www.evaluationcanada.ca/txt/1_rfp_e.pdf

Canadian Evaluation Society National Council. (2007a). Letter to CES member. Retrieved November 26, 2007, from http://www.evaluationcanada.ca/site.cgi?s=5&ss=6&_lang=EN

Canadian Evaluation Society National Council. (2007b). Professional designation project plan. Retrieved April 1, 2008, from http://www.evaluationcanada.ca/txt/profdes_plan_e.pdf

Canadian Evaluation Society National Council. (2007c). Response to proposed: Action plan for the Canadian Evaluation Society with respect to professional standards for evaluators. Retrieved November 17, 2007, from http://www.evaluationcanada.ca/site.cgi?s=5&ss=6&_lang=en

Christie, C. A. (2003). What guides evaluation? A study of how evaluation practice maps onto evaluation theory. *New Directions for Program Evaluation, 97*, 7-35.

Connell, M. W., Sheridan, K., & Gardner, H. (2003). On abilities and domains. In R. J. Sternberg & E. L. Grigorenko (Eds.), *The psychology of abilities, competencies, and expertise* (pp. 126-155). Cambridge, UK: Cambridge University Press.

Cousins, J. B., Buchanan, H., Kuji-Shikatani, K., & Maicher, B. (2008, November 8). *Toward professional designations for evaluators: The Canadian experience.* Paper presented at the Annual American Evaluation Association Conference, Denver, CO.

Cousins, J. B., Cullen, J., Malik, S., & Maicher, B. (2007). *Warming up to the prospect of professional designations: Reflections on the Canadian process.* Paper presented at the Annual American Evaluation Association Conference, Baltimore, MD.

Cousins, J. B., & Earl, L. M. (1999). When the boat gets missed: Response to M. F. Smith. *American Journal of Evaluation, 20*(2), 309-317.

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin, 52*(4), 281-302.

Dillman, D. A. (2007). *Mail and internet surveys: The tailored design method* (2nd ed.). Hoboken, N.J.: John Wiley & Sons, Inc.

Epstein, R. M., & Hundert, E. M. (2002). Defining and Assessing Professional Competence. *JAMA, 287*(2), 226-235.

Equal Employment Opportunity Commission, Civil Service Commission, Department of Labor, & Department of Justice. (1978). Uniform guidelines on employee selection procedures. *Federal Register, 43*(166), 38295-38309.

Eraut, M. P. (1994). *Developing professional knowledge and competence*. London: Routledge.

Evers, F. T., Rush, J. C., & Berdrow, I. (1998). *The base of competence: Skills for lifelong learning and employability*. San Francisco, CA: Jossey-Bass.

Fitzpatrick, J. L., Sanders, J. R., & Worthen, B. R. (2004). *Program evaluation: Alternative approaches and practical guidelines* (3rd ed.). Boston: Pearson.

Flanagan, J. C. (1954). The critical incident technique. *Psychological Bulletin, 51*(4), 327-358.

Fleishman, E. A. (1953). The description of supervisory behavior. *Journal of Applied Psychology, 37*(1), 1-6.

Ghere, G., King, J. A., Stevahn, L., & Minnema, J. (2006). A professional development unit for reflecting on program evaluator competencies. *American Journal of Evaluation, 27*(1), 108-123.

Green, P. C. (1999). *Building robust competencies: Linking human resource systems to organizational strategies*. San Francisco, CA: Jossey-Bass.

Gussman, T. K. (2005). *Improving the professionalism of evaluation*. Ottawa: Centre for Excellence in Evaluation, Treasury Board Secretariat of Canada.

Halpern, G., Gauthier, B., & McDavid, J. C. (2007). An action plan for the Canadian Evaluation Society with respect to professional standards for evaluators. Retrieved April 1, 2008, from http://www.evaluationcanada.ca/txt/4_consortium_e.pdf

Halpern, G., & Long, B. (2007). Interview results: Professional designations for evaluators. Retrieved April 1, 2008, from http://www.evaluationcanada.ca/site.cgi?s=5&ss=6&_lang=EN

Hoffmann, T. (1999). The meanings of competency. *Journal of European Industrial Training, 23*(6), 275-285.

Hoge, M. A., Tondora, J., & Marrelli, A. F. (2005). The fundamentals of workforce competency: Implications for behavioral health. *Administration and Policy in Mental Health and Mental Health Services Research, 32*(5/6), 509-531.

Horton, S. (2000). Introduction: The competency movement: Its origins and impact on the public sector. *The International Journal of Public Sector Management, 13*(4), 306-318.

Howell, W. S. (1982). *The empathic communicator*. Belmont, CA: Wadsworth Publishing Company.

Huse, I., & McDavid, J. C. (2006). Literature review: Professionalization of evaluators. Retrieved November 16, 2007, from http://www.evaluationcanada.ca/txt/2_literature_e.pdf

Johnson Jr., H. C. (1984). Teacher competence: A history analysis. In E. C. Short (Ed.), *Competence: Inquiries into its meaning and acquisition in educational settings* (pp. 41-70). Lanham, MD: University Press of America.

Joint Committee on Standards for Educational Evaluation. (1981). *Standards for evaluations of educational programs, projects, and materials*. New York: McGraw-Hill.

Joint Committee on Standards for Educational Evaluation. (1994). *The program evaluation standards* (2nd ed.). Thousand Oaks, CA: Sage.

Jones, S. C., & Worthen, B. R. (1999). AEA members' opinions concerning evaluator certification. *American Journal of Evaluation, 20*(3), 495-506.

Kane, M. T. (2001). Current concerns in validity theory. *Journal of Educational Measurement, 38*(4), 319-342.

King, J. A., Stevahn, L., Ghere, G., & Minnema, J. (2001). Toward a taxonomy of essential evaluator competencies. *American Journal of Evaluation, 22*(2), 229-247.

Kirkhart, K. (1981). Defining evaluator competencies: New light on an old issue. *American Journal of Evaluation, 2*, 188-192.

Kliebart, H. M. (1995). *The struggle for the American curriculum, 1893-1958* (2nd ed.). New York: Routledge.

Leonard, D., & Sensiper, S. (1998). The role of tacit knowledge in group innovation. *California Management Review, 40*(3), 112-132.

Loevinger, J. (1957). Objective tests as instruments of psychological theory. *Psychological Reports, 3*, 635-694.

Long, B. (2007). Minority report with respect to professional designations for evaluators. Retrieved April 1, 2008, from http://www.evaluationcanada.ca/txt/5_minority.pdf

Long, B., & Kishchuk, N. (1997). *Professional certification: A report to the National Council of the Canadian Evaluation Society on the experience of other organizations*. Ottawa: Canadian Evaluation Society.

Love, A. J. (1994). Should evaluators be certified? *New Directions for Program Evaluation, 62*, 29-40.

Maicher, B. (2008, November 6). *A competency profile for Canadian evaluators.* Paper presented at the Annual American Evaluation Association Conference, Denver, CO.

McCain, M. (2002). National skill standards board: Integrating manufacturing skill standards into education and training programs. Retrieved February 26, 2008, from http://www.aypf.org/forumbriefs/2002/fb020102.htm

McClelland, D. C. (1973). Testing for competence rather than intelligence. *American Psychologist, 28*(1), 1-14.

McClelland, D. C. (1993). Introduction. In L. M. Spencer & S. M. Spencer (Eds.), *Competence at work: Models for superior performance*. New York: Wiley.

McLagan, P. A. (1997). Competencies: The next generation. *Training and Development, 51*(5), 40-47.

Mertens, D. M. (1994). Training evaluators: Unique skills and knowledge. *New Directions for Program Evaluation, 62*, 17-28.

Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13-103): Macmillan Publishing Company.

Messick, S. (1995a). Standards of validity and the validity of standards in performance assessment. *Educational Measurement: Issues and Practice, 14*(4), 5-8.

Messick, S. (1995b). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist, 50*(9), 741-749.

National Skill Standards Act of 1994, Sec. 502. (1994).

Neumann, W. (1979). Educational responses to the concern for proficiency. In G. Grant, P. Elbow, T. Ewens, Z. Gamson, W. Kohli, W. Neumann, V. Olesen & D. Riesman (Eds.), *On competence: A critical analysis of competence-based reforms in higher education* (pp. 66-94). San Francisco, CA: Jossey-Bass.

OECD. (2002). Definition and selection of competencies (DeSeCo): Theoretical and conceptual foundations. *Strategy paper JT00132752* Retrieved January 7, 2008, from http://www.portal-stat.admin.ch/deseco/deseco_strategy_paper_final.pdf

Orcher, L. T. (2005). *Conducting research: Social and behavioral science methods*. Glendale, CA: Pyrczak Publishing.

Patten, M. L. (2005). *Proposing empirical research: A guide to the fundamentals* (3rd ed.). Glendale, CA: Pyrczak Publishing.

Patton, M. Q. (1990). The challenge of being a profession. *Evaluation Practice, 11*(1), 45-51.

Polanyi, M. (1967). *The Tacit Dimension*. London: Routledge & Kegan Paul.

Roe, R. A. (1999). Work performance: A multiple regulation perspective. In C. L. Cooper & I. T. Robertson (Eds.), *International review of industrial and organizational psychology* (Vol. 14, pp. 231-336). Chichester, NY: Wiley.

Roe, R. A. (2002). What makes a competent psychologist? *European Psychologist, 7*(3), 192-203.

Roe, R. A. (2004, March 26). *Towards a competence based curriculum for European W&O psychology.* Paper presented at the Annual European Network of Work & Organizational Psychologists (ENOP) Symposia, Paris.

Rychen, D. S. (2001). Introduction. In D. S. Rychen & L. H. Salganik (Eds.), *Defining and selecting key competencies*. Kirkland, WA: Hogrefe & Huber Publishers.

Rychen, D. S. (2004). Key competencies for all: An overarching conceptual frame of reference. In D. S. Rychen & A. Tiana (Eds.), *Developing key competencies in education: Some lessons from international and national experience*. France: UNESCO: International Bureau of Education.

Salganik, L. H. (2001). Competencies for life: A conceptual and empirical challenge. In D. S. Rychen & L. H. Salganik (Eds.), *Defining and selecting key competencies* (pp. 17-32). Kirkland, WA: Hogrefe & Huber Publishers.

Schoonover Associates. (2003). Competency FAQ's. Retrieved February 18, 2008, from http://www.schoonover.com/competency_faqs.htm

Schwandt, T. A. (2008, November 8). *Remarks on panel on "Competencies, Credentials, Certification, and Professional Development: Experience & Issues for Evaluation Policy"*. Paper presented at the Annual American Evaluation Association Conference, Denver, CO.

Scriven, M. (1996). Types of evaluation and types of evaluator. *Evaluation Practice, 17*(2), 151-161.

Shadish, W. R., Cook, T. D., & Leviton, L. C. (1991). *Foundations of program evaluation: Theories of practice*. Thousand Oaks, CA: Sage.

Shippmann, J. S., Ash, R. A., Battista, M., Carr, L., Eyde, L. D., Hesketh, B., et al. (2000). The practice of competency modeling. *Personnel Psychology, 53*(3), 703-740.

Smith, M. F. (1999). Should AEA begin a process for restricting membership in the profession of evaluation? *American Journal of Evaluation, 20*(3), 521-531.

Smith, M. K. (1996). Competence and competency. Retrieved November 7, 2007, from http://www.infed.org/biblio/b-comp.htm

Spencer, L. M., & Spencer, S. M. (1993). *Competence at work: Models for superior performance*. New York: Wiley.

Stevahn, L., King, J. A., Ghere, G., & Minnema, J. (2005). Establishing essential competencies for program evaluators. *American Journal of Evaluation, 26*(1), 43-59.

Swinton, L. *Smooth your learning journey with the learning matrix*. Retrieved April 1, 2008, from http://www.mftrou.com/support-files/learning-matrix.pdf.

Taylor, F. W. (1911). *The principles of scientific management*. Retrieved December 8, 2007, from http://www.marxists.org/reference/subject/economics/taylor/principles/index.htm.

Thorndike, R. M. (2005). *Measurement and evaluation in psychology and education*. Upper Saddle River, NJ: Pearson Merrill Prentice Hall.

Toal, S. A. (2007). *The development and validation of an Evaluation Involvement Scale for Use in multi-site evaluations* (Doctoral dissertation). University of Minnesota, Minneapolis, MN.

Tuxworth, E. (1989). Competence based education and training: Background and origins. In J. W. Burke (Ed.), *Competency based education and training* (pp. 10-25). Bristol, PA: The Falmer Press.

Weinert, F. E. (2001). Concept of competence: A conceptual clarification. In D. S. Rychen & L. H. Salganik (Eds.), *Defining and selecting key competencies* (pp. 45-66). Kirkland, WA: Hogrefe & Huber Publishers.

Weiss, C. H. (1987). Evaluating Social Programs: What have we learned? *Society, 25*(1), 40-45.

What is a web survey? (2003). Retrieved June 8, 2009, from http://www.activecampaign.com/blog/2009/02/03/what-is-a-web-survey/

Wholey, J. S. (1986). Using Evaluation to Improve Government Performance. *American Journal of Evaluation, 7*(2), 5-13.

Willis, G. B. (2004). *Cognitive interviewing: A tool for improving questionnaire design*. Thousand Oaks, CA: Sage Publications.

Worthen, B. R. (1994). Is evaluation a mature profession that warrants the preparation of evaluation professionals? *New Directions for Program Evaluation, 62*, 3-15.

Worthen, B. R. (1999). Critical challenges confronting certification of evaluators. *American Journal of Evaluation, 20*(3), 533-555.

Zorzi, R., McGuire, M., & Perrin, B. (2002). Evaluation benefits, outputs, and knowledge elements: Canadian Evaluation Society project in support of advocacy and professional development. Retrieved March 27, 2008, from http://consultation.evaluationcanada.ca/pdf/ZorziCESReport.pdf

Appendix A

Essential Competencies for Program Evaluators Survey

The Essential Competencies for Program Evaluators (ECPE) is a taxonomy originally developed in 2001 that lists knowledge, skills, and dispositions for program evaluators. This survey seeks your input on the extent to which each competency is necessary. The wording of each is exactly as it appears in the ECPE.

Click on the radio button on this 5-point scale to indicate the extent to which you believe that program evaluators need to be able to demonstrate this competency, whether or not you do so regularly.					
	Do not think about your own experience. Think about evaluators in general.				
	Not at all necessary	Slightly necessary	Moderately necessary	Strongly necessary	Extremely necessary
1.1 Applies professional evaluation standards	○	○	○	○	○
1.2 Acts ethically and strives for integrity and honesty in conducting evaluations	○	○	○	○	○
1.3 Conveys personal evaluation approaches and skills to potential clients	○	○	○	○	○
1.4 Respects clients, respondents, program participants, and other stakeholders	○	○	○	○	○
1.5 Considers the general and public welfare in evaluation practice	○	○	○	○	○
1.6 Contributes to the knowledge base of evaluation	○	○	○	○	○
Input comments you have on any of the above competencies (for example, explanations, changes, revisions, deletions). Please refer to each by number					

Click on the radio button on this 5-point scale to indicate the extent to which you believe that program evaluators need to be able to demonstrate this competency, whether or not you do so regularly.					
	Do not think about your own experience. Think about evaluators in general.				
	Not at all necessary	Slightly necessary	Moderately necessary	Strongly necessary	Extremely necessary
2.1 Understands the knowledge base of evaluation (terms, concepts, theories, assumptions)	○	○	○	○	○
2.2 Knowledgeable about quantitative methods	○	○	○	○	○
2.3 Knowledgeable about qualitative methods	○	○	○	○	○
2.4 Knowledgeable about mixed methods	○	○	○	○	○
2.5 Conducts literature reviews	○	○	○	○	○
Input comments you have on any of the above competencies (for example, explanations, changes, revisions, deletions). Please refer to each by number					

Click on the radio button on this 5-point scale to indicate the extent to which you believe that program evaluators need to be able to demonstrate this competency, whether or not you do so regularly.					
	Do not think about your own experience. Think about evaluators in general.				
	Not at all necessary	Slightly necessary	Moderately necessary	Strongly necessary	Extremely necessary
2.6 Specifies program theory	○	○	○	○	○
2.7 Frames evaluation questions	○	○	○	○	○
2.8 Develops evaluation designs	○	○	○	○	○
2.9 Identifies data sources	○	○	○	○	○
2.10 Collects data	○	○	○	○	○
Input comments you have on any of the above competencies (for example, explanations, changes, revisions, deletions). Please refer to each by number					

Click on the radio button on this 5-point scale to indicate the extent to which you believe that program evaluators need to be able to demonstrate this competency, whether or not you do so regularly.					
	Do not think about your own experience. Think about evaluators in general.				
	Not at all necessary	Slightly necessary	Moderately necessary	Strongly necessary	Extremely necessary
2.11 Assesses validity of data	○	○	○	○	○
2.12 Assesses reliability of data	○	○	○	○	○
2.13 Analyzes data	○	○	○	○	○
2.14 Interprets data	○	○	○	○	○
2.15 Makes judgments	○	○	○	○	○
Input comments you have on any of the above competencies (for example, explanations, changes, revisions, deletions). Please refer to each by number					

Click on the radio button on this 5-point scale to indicate the extent to which you believe that program evaluators need to be able to demonstrate this competency, whether or not you do so regularly.					
	Do not think about your own experience. Think about evaluators in general.				
	Not at all necessary	Slightly necessary	Moderately necessary	Strongly necessary	Extremely necessary
2.16 Develops recommendations	○	○	○	○	○
2.17 Provides rationales for decisions throughout the evaluation	○	○	○	○	○
2.18 Reports evaluation procedures and results	○	○	○	○	○
2.19 Notes strengths and limitations of the evaluation	○	○	○	○	○
2.20 Conducts meta-evaluations	○	○	○	○	○
Input comments you have on any of the above competencies (for example, explanations, changes, revisions, deletions). Please refer to each by number					

Click on the radio button on this 5-point scale to indicate the extent to which you believe that program evaluators need to be able to demonstrate this competency, whether or not you do so regularly.					
	Do not think about your own experience. Think about evaluators in general.				
	Not at all necessary	Slightly necessary	Moderately necessary	Strongly necessary	Extremely necessary
3.1 Describes the program	○	○	○	○	○
3.2 Determines program evaluability	○	○	○	○	○
3.3 Identifies the interests of relevant stakeholders	○	○	○	○	○
3.4 Serves the information needs of intended users	○	○	○	○	○
3.5 Addresses conflicts	○	○	○	○	○
3.6 Examines the organizational context of the evaluation	○	○	○	○	○
Input comments you have on any of the above competencies (for example, explanations, changes, revisions, deletions). Please refer to each by number					

Click on the radio button on this 5-point scale to indicate the extent to which you believe that program evaluators need to be able to demonstrate this competency, whether or not you do so regularly.					
	Do not think about your own experience. Think about evaluators in general.				
	Not at all necessary	Slightly necessary	Moderately necessary	Strongly necessary	Extremely necessary
3.7 Analyzes the political considerations relevant to the evaluation	○	○	○	○	○
3.8 Attends to issues of evaluation use	○	○	○	○	○
3.9 Attends to issues of organizational change	○	○	○	○	○
3.10 Respects the uniqueness of the evaluation site and client	○	○	○	○	○
3.11 Remains open to input from others	○	○	○	○	○
3.12 Modifies the study as needed	○	○	○	○	○
Input comments you have on any of the above competencies (for example, explanations, changes, revisions, deletions). Please refer to each by number					

Click on the radio button on this 5-point scale to indicate the extent to which you believe that program evaluators need to be able to demonstrate this competency, whether or not you do so regularly.					
	Do not think about your own experience. Think about evaluators in general.				
	Not at all necessary	Slightly necessary	Moderately necessary	Strongly necessary	Extremely necessary
4.1 Responds to requests for proposals	○	○	○	○	○
4.2 Negotiates with clients before the evaluation begins	○	○	○	○	○
4.3 Writes formal agreements	○	○	○	○	○
4.4 Communicates with clients throughout the evaluation process	○	○	○	○	○
4.5 Budgets an evaluation	○	○	○	○	○
4.6 Justifies cost given information needs	○	○	○	○	○
Input comments you have on any of the above competencies (for example, explanations, changes, revisions, deletions). Please refer to each by number					

Click on the radio button on this 5-point scale to indicate the extent to which you believe that program evaluators need to be able to demonstrate this competency, whether or not you do so regularly.					
	Do not think about your own experience. Think about evaluators in general.				
	Not at all necessary	Slightly necessary	Moderately necessary	Strongly necessary	Extremely necessary
4.7 Identifies needed resources for evaluation, such as information, expertise, personnel, instruments	○	○	○	○	○
4.8 Uses appropriate technology	○	○	○	○	○
4.9 Supervises others involved in conducting the evaluation	○	○	○	○	○
4.10 Trains others involved in conducting the evaluation	○	○	○	○	○
4.11 Conducts the evaluation in a nondisruptive manner	○	○	○	○	○
4.12 Presents work in a timely manner	○	○	○	○	○
Input comments you have on any of the above competencies (for example, explanations, changes, revisions, deletions). Please refer to each by number					

Click on the radio button on this 5-point scale to indicate the extent to which you believe that program evaluators need to be able to demonstrate this competency, whether or not you do so regularly.					
	Do not think about your own experience. Think about evaluators in general.				
	Not at all necessary	Slightly necessary	Moderately necessary	Strongly necessary	Extremely necessary
5.1 Aware of self as an evaluator (knowledge, skills, dispositions)	○	○	○	○	○
5.2 Reflects on personal evaluation practice (competencies and areas for growth)	○	○	○	○	○
5.3 Pursues professional development in evaluation	○	○	○	○	○
5.4 Pursues professional development in relevant content areas	○	○	○	○	○
5.5 Builds professional relationships to enhance evaluation practice	○	○	○	○	○
Input comments you have on any of the above competencies (for example, explanations, changes, revisions, deletions). Please refer to each by number					

Click on the radio button on this 5-point scale to indicate the extent to which you believe that program evaluators need to be able to demonstrate this competency, whether or not you do so regularly.					
	Do not think about your own experience. Think about evaluators in general.				
	Not at all necessary	Slightly necessary	Moderately necessary	Strongly necessary	Extremely necessary
6.1 Uses written communication skills	○	○	○	○	○
6.2 Uses verbal/listening communication skills	○	○	○	○	○
6.3 Uses negotiation skills	○	○	○	○	○
6.4 Uses conflict resolution skills	○	○	○	○	○
6.5 Facilitates constructive interpersonal interaction (teamwork, group facilitation, processing)	○	○	○	○	○
6.6 Demonstrates cross-cultural competence	○	○	○	○	○
Input comments you have on any of the above competencies (for example, explanations, changes, revisions, deletions). Please refer to each by number					

On the next page, please answer some questions regarding your experience with program evaluation.

What is your job title? Select *one*.
○ Evaluation consultant
○ Evaluation director
○ Evaluation staff
○ Program director
○ Program staff
○ Other, please specify: ________________________________

Are you a graduate student?
○ Yes
○ No

In what types of organizations do you work? Check *all* that apply.
○ K-12 education/school district
○ College/university
○ Research firm
○ Nonprofit organization
○ Business/industry
○ Health
○ Government
○ Other, please specify: ________________________________

Are you self-employed?
○ Yes
○ No

How many years have you been involved in activities related to program evaluation, for example, planning, data collection, data analysis? Select *one*.
○ <1
○ 2
○ 3
○ 4
○ 5
○ 6
○ 7
○ 8
○ 9
○ 10
○ 11
○ 12
○ 13

○ 14
○ >15

How would you label yourself as an evaluator? Select *one*.
○ Novice/entry
○ Proficient
○ Skilled
○ Expert/mastery

If your highest earned degree is a doctoral degree, what field is your doctorate in?
○ Doctoral degree in program evaluation/evaluation studies
○ Doctoral degree in research design/methods
○ Doctoral degree in another discipline; please specify: ___________________________

If your highest earned degree is a master's degree, what field is your master's in?
○ Master's degree in program evaluation/evaluation studies
○ Master's degree in research design/methods
○ Master's degree in another discipline; please specify: ___________________________

Have you earned a graduate certificate in program evaluation?
○ Yes
○ No

What types of training related to program evaluation have you completed? Check *all* that apply.
○ Evaluation theory
○ Evaluation practice (how to do evaluations)
○ Research methods/design
○ Qualitative methods
○ Quantitative methods
○ Mixed methods
○ Measurement
○ Other, please specify: ________________________________

What's your gender?
○ Male
○ Female

What is the state of your residence? _______________________

Do you belong to one or more professional evaluation associations? If yes, please list them.

If you have any other comments about the Essential Competencies for Program Evaluators, please add them here.

Thank you for completing this survey!
You can now close this window.

Appendix B

Interview Protocol

- Get oral consent
- Present the list of Essential Competencies for Program Evaluators to the interviewee.
- Talk about the study and the 61 competencies.
- Ask the general question:
    - In general, how would you use the ECPE in your field?
- Then guide the interviewee to look at one category at a time.
- Ask the following three questions for each category.

Professional Practice

1. Do you think these six competencies actually measure an evaluator's professional ability?
2. What are some additions, deletions, or other changes you think necessary to this category?
3. How would you use these six professional practice competencies in your field?

Systematic Inquiry

1. Do you think these twenty competencies actually measure an evaluator's competence to make systematic inquiry?
2. What are some additions, deletions, or other changes you think necessary to this category?
3. How would you use these twenty systematic inquiry competencies in your field?

Situational Analysis

1. Do you think these twelve competencies actually measure an evaluator's competence to conduct situational analysis?
2. What are some additions, deletions, or other changes you think necessary to this category?
3. How would you use these twelve situational analysis competencies in your field?

Project Management

1. Do you think these twelve competencies actually measure an evaluator's ability to manage the evaluation project?
2. What are some additions, deletions, or other changes you think necessary to this category?
3. How would you use these twelve project management competencies in your field?

Reflective Practice

1. Do you think these five competencies actually measure an evaluator's reflective practice?

2.What are some additions, deletions, or other changes you think necessary to this category?
3.How would you use these five reflective practice competencies in your field?

Interpersonal Competence

1.Do you think these six competencies actually measure an evaluator's interpersonal competence?
2.What are some additions, deletions, or other changes you think necessary to this category?
3.How would you use these six interpersonal competencies in your field?

- At the end, ask a couple more questions:
  - Is there any other field that you can think of that ECPE would be useful in?
  - Were the ECPE to be institutionalized in the form of degree programs, licensure, certification, and accreditation, what do you think would be the positive and negative consequences?
- Do you have any other comments about the ECPE?