

5-2014

# Metaevaluation of a university teacher education assessment system.

Katherine Amelia Shanahan  
*University of Louisville*

Follow this and additional works at: <https://ir.library.louisville.edu/etd>

Part of the [Education Commons](#)

---

## Recommended Citation

Shanahan, Katherine Amelia, "Metaevaluation of a university teacher education assessment system." (2014). *Electronic Theses and Dissertations*. Paper 1311.  
<https://doi.org/10.18297/etd/1311>

This Doctoral Dissertation is brought to you for free and open access by ThinkIR: The University of Louisville's Institutional Repository. It has been accepted for inclusion in Electronic Theses and Dissertations by an authorized administrator of ThinkIR: The University of Louisville's Institutional Repository. This title appears here courtesy of the author, who has retained all other copyrights. For more information, please contact [thinkir@louisville.edu](mailto:thinkir@louisville.edu).

METAEVALUATION OF A UNIVERSITY TEACHER EDUCATION  
ASSESSMENT SYSTEM

By

Katherine Amelia Shanahan  
B.A., DePauw University, 2003  
M.Ed., University of Louisville, 2009

A Dissertation  
Submitted to the Faculty of the  
College of Education and Human Development of the University of Louisville  
in Partial Fulfillment of the Requirements  
for the Degree of

Doctor of Philosophy

Leadership, Foundations, and Human Resource Education  
University of Louisville  
Louisville, Kentucky

May 2014

Copyright 2013 by Katherine A. Shanahan

All rights reserved



METAEVALUATION OF A UNIVERSITY TEACHER EDUCATION  
ASSESSMENT SYSTEM

By

Katherine Amelia Shanahan  
B.A., DePauw University, 2003  
M.Ed., University of Louisville, 2009

A Dissertation Approved on

March 25, 2014

by the following Dissertation Committee:

---

Dr. Namok Choi, Chair

---

Dr. Ann Larson

---

Dr. Donna Gaus

---

Robert Goldstein

## DEDICATION

This dissertation is dedicated to my parents and husband

Don and Nancy Hunt, and David Shanahan

for your infinite love and support.

## ACKNOWLEDGEMENTS

I would like to express deepest gratitude to Dr. Namok Choi who served as my dissertation committee chair. Thank you Dr. Choi for your guidance, support, and investment in my academic and professional endeavors. I will cherish the opportunity I have had to work with you. I would also like to thank my committee members, Dr. Ann Larson, Dr. Donna Gaus, and Vice Provost Robert Goldstein for their commitment and for providing their time and expertise to refine and develop this study. I am honored to have each of you serve on this committee. A special thank you to Dr. Cheryl Gilchrist and Joanne Webb for contributions to this study and collaboration in determining next steps of the metaevaluation process. Additionally, I would like to thank Dr. Jill Adelson, Dr. Joseph Petrosko, Dr. Blake Haselton, Dr. Jake Gross, and Dr. Jeffrey Valentine who were all instrumental in my academic and professional development.

To my colleagues in the Office of Academic Affairs and Unit Effectiveness, Ann Larson, Joanne Webb, Robin Dorsey, Linda Moore, Melissa Lankford, Dhiane Bradley, and Elise Romines, thank you for your friendship and support throughout this process. I am truly blessed to have such outstanding colleagues who support and encourage me on a daily basis. A special thank you to Vice Dean Ann Larson and Dean Haselton for your leadership, encouragement, and support in my academic pursuits. I would like to extend my appreciation to the many close friends and colleagues that I have had the opportunity

to work with at the University of Louisville and across the state of Kentucky. I have greatly valued the opportunities to collaborate and learn from my peers in K-ITEP.

This dissertation is possible because of the support and the love of my family. My parents commitment to God, family, and community have inspired me and encouraged me to make a positive difference in this world. I am so fortunate to be surrounded by such a loving and caring family. To my grandparents, thank you for the family you have raised and the values that you have instilled in all of us. Thank you to my brother, who is an inspiration in all that he does.

A special thank you to my amazing husband, David. Thank you for encouraging me to pursue my education and advancement. Thank you for the many times you have put everything on hold to allow me to reach my goals and for holding my hand through this process. I am so grateful for your love and support throughout this process.



## ABSTRACT

### METAEVALUATION OF A UNIVERSITY TEACHER EDUCATION ASSESSMENT SYSTEM

Katherine A. Shanahan

March 25, 2014

Metaevaluation is the evaluation of an evaluation or evaluation system (Scriven, 1969). It serves as a mechanism to ensure quality in evaluation approaches and implementation. Operationally metaevaluation is defined as “the process of delineating, obtaining, and applying descriptive information and judgmental information – about the utility, feasibility, propriety, and accuracy of an evaluation and its systematic nature, competent conduct, integrity/honesty, respectfulness, and social responsibility to guide the evaluation and/or report its strengths and weaknesses” (Stufflebeam, 2001, *p.*185).

This study was a metaevaluation of an assessment system designed for accreditation requirements to support continuous improvement in teacher education programs at the University of Louisville. The study was intended to serve as a formative metaevaluation to identify strengths and weaknesses in the University of Louisville, College of Education and Human Development’s (CEHD) teacher education assessment system to support improvement of the system and better support continuous improvement of teacher education programs. The study took careful consideration of accountability and accreditation requirements, as well as evaluation and metaevaluation standards and practices. The study utilized Stufflebeam’s structure for metaevaluation (2001), which

supports strategic and contextual analysis of the evaluation or evaluation system to address alignment with stakeholders needs.

The study employed mixed methods to address four research questions. The research questions were focused on the application of data from the CEHD's assessment system in driving program improvement and also the reliability and validity of instruments used in the assessment system.

The first research question was focused on identifying the types of assessments that best support program improvement in teacher education. A qualitative case study analysis revealed a lack of explicit connections to data within the CEHD's SLO action plans in which faculty identify plans for improving programs. Implied connections to data, included references to the 10 Unit Key Assessments, Hallmark Assessment Tasks (HATs), and indirect assessment data (QMS student satisfaction survey data. These results indicate that a variety of assessments support program improvement and are in alignment with CAEP standards (2013), the American Evaluation Association (2013), and the Joint Committee on Standards for Educational Evaluation (2011), multiple measures are necessary in sound evaluation and evaluation systems. This study resulted in recommendations to modify SLO templates and action plan prompts to ensure more explicit connections of data to the action plans and even follow-through on action plans.

The second research question was intended to identify how assessment data are used to drive continuous improvement in teacher education programs. The qualitative case study review of SLO action plans and reflections on previous year's plans for improvement identified actions in the area of curriculum, faculty development, assessments, field and clinical experiences, and candidate performance. These findings

demonstrated a real strength of the CEHD's assessment system, as it demonstrates that the assessment system is driving continuous program improvement. One suggestion for improvement was increased documentation related to follow-through of actions within the current assessment system structures.

The third research question pertained to reliability of instruments used across programs. The analysis revealed no concerns in regards to reliability of instruments across programs. The CEHD is encouraged to incorporate continued training and collaborative sessions to dissect and practice application of instruments to ensure reliability over time. This is especially important as programs revise instruments, assessors matriculate, and assessment context changes.

The fourth and final research questions reviewed the construct validity of instruments in the CEHD assessment system aligned with the CEHD's conceptual framework. The study revealed adequate construct validity related to measuring critical thinking, problem solving, and professional leadership, however also revealed potential concerns regarding discriminant validity. To address these findings, it has been recommended that the CEHD transition to 4-point rubrics instead of the current 3-point rubrics used in the assessment system. The study has outlined next steps in making that transition.

In conclusion, this study identified strengths in the reliability of instrumentation and strategic application of data. Areas for improvement include revision of instruments to provide differentiation between performance levels and outcomes in the assessment system and revisions to SLO processes and templates to ensure more explicit connections between data and decision making. Ultimately, this metaevaluation has identified the

most pertinent next steps for CEHD administrators, faculty, and staff in improving the assessment system to drive continuous program improvement in alignment with the Council for the Accreditation of Educator Preparation (CAEP) and the Kentucky Education Professional Standards Board (EPSB) accreditation processes.

## TABLE OF CONTENTS

|  | PAGE |
|--|------|
| ACKNOWLEDGEMENTS.....                                  | iv   |
| ABSTRACT.....  | vi   |
| LIST OF TABLES.....                                    | xiv  |
| LIST OF FIGURES.....                                   | xvii |
| I. INTRODUCTION.....                                   | 1    |
| University of Louisville .....                         | 2    |
| College of Education and Human Development (CEHD)..... | 3    |
| CEHD Conceptual Framework.....                         | 4    |
| CEHD Accreditation.....                                | 6    |
| Assessment System.....                                 | 7    |
| Purpose of Study.....                                  | 13   |
| Research Questions.....                                | 16   |
| Limitations.....                                       | 17   |
| Significance of the Study.....                         | 19   |
| Definitions.....                                       | 21   |
| II. LITERATURE REVIEW.....                             | 23   |
| Accountability in Higher Education.....                | 23   |
| Public Accountability.....                             | 29   |
| Government Accountability.....                         | 32   |

|   |    |
|---|----|
| Accreditation.....  | 35 |
| Evaluation.....   | 38 |
| Evaluation Approaches.....  | 39 |
| Evaluation in Education.....  | 42 |
| Formative and Summative Evaluation.....   | 44 |
| Internal and External Evaluation.....   | 46 |
| Evaluation Standards, Guiding Principles, and Checklists.....                     | 49 |
| Metaevaluation.....   | 51 |
| Metaevaluation Approach.....  | 53 |
| Evaluation versus Metaevaluation.....   | 57 |
| Critical Thinking, Problem Solving, and Professional Leadership.....              | 58 |
| Inquiry and Critical Thinking.....  | 60 |
| Action and Problem Solving.....   | 62 |
| Advocacy and Professional Leadership.....   | 63 |
| Critical Thinking, Problem-Solving, and Professional Leadership<br>Framework..... | 65 |
| III. METHODS.....   | 67 |
| Stakeholders.....   | 67 |
| Metaevaluator.....  | 71 |
| Metaevaluation Questions (Research Questions).....                                | 72 |
| Alignment to Metaevaluation Standards.....  | 77 |
| Written Agreement.....  | 78 |
| Collecting Information.....   | 79 |

|  |     |
|--|-----|
| Methods.....   | 79  |
| Participants.....  | 86  |
| Analyzing Findings and Judging Adherence to Standards.....                           | 89  |
| IV. RESULTS.....   | 91  |
| Data.....  | 92  |
| Descriptive Statistics for Quantitative Analysis.....                                | 93  |
| Qualitative Case Study Analysis.....   | 94  |
| Assessments Used to Support Program Improvement.....                                 | 96  |
| Application of Assessment Findings.....  | 100 |
| Reliability Analysis.....  | 105 |
| Validity Analysis.....   | 117 |
| Summary of Results.....  | 136 |
| V. DISCUSSION.....   | 139 |
| Assessments Driving Program Improvement.....   | 140 |
| Impact of the Assessment System.....   | 142 |
| Reliability of the Assessment System.....  | 143 |
| Validity of the Assessment System.....   | 144 |
| Alignment to CAEP Standards.....   | 147 |
| Adherence to JCSEE Standards.....  | 153 |
| Recommendations for Next Steps.....  | 157 |
| Revision of Prompts in Action Plans.....   | 159 |
| Increased Emphasis on the Reflection on Previous Year’s Plan for<br>Improvement..... | 160 |

|   |     |
|---|-----|
| Revision of Format for Student Learning Outcome Reports.....                                    | 161 |
| Revision to Incorporation of Feedback in SLO Process.....                                       | 162 |
| Revision of Rubrics.....  | 163 |
| Faculty Development.....  | 164 |
| Implications for Teacher Education Programs and Evaluation Field.....                           | 165 |
| Limitations.....  | 166 |
| Study Conclusions.....  | 167 |
| REFERENCES.....   | 170 |
| APPENDIX A: Sample Continuous Assessment Record and Documentation System<br>(CARDS) Chart ..... | 179 |
| APPENDIX B: JCSEE Program Evaluation Standards.....   | 184 |
| APPENDIX C: One-Way Random Effects ANOVA Results.....   | 187 |
| APPENDIX D: Institutional Review Board.....   | 198 |
| APPENDIX E: Addendum to IRB Study.....  | 209 |
| CURRICULUM VITAE.....   | 210 |



## LIST OF TABLES

| TABLE   | PAGE |
|---|------|
| 1. CEHD Conceptual Framework Aligned with Candidate Knowledge, Skills, and Dispositions (2007).....       | 5    |
| 2. 10 Unit Key Assessments for Educator Preparation.....  | 9    |
| 3. CAEP Standards.....  | 14   |
| 4. Eight Premises for Meta-Evaluation (Stufflebeam, 2011).....  | 52   |
| 5. Stufflebeam’s Structure for Identifying Alternative Metaevaluation Procedures (Stufflebeam, 2001)..... | 53   |
| 6. CEHD Conceptual Framework Aligned with Candidate Knowledge, Skills, and Dispositions (2007).....       | 59   |
| 7. Research Questions.....  | 73   |
| 8. Milestone Measures for CEHD Rubrics.....   | 74   |
| 9. Model Constructs, Measured Items, and Item Description.....  | 76   |
| 10. Metaevaluation Checklist Standards (Stufflebeam, 1999).....   | 77   |
| 11. Initial Certification Student Learning Outcomes (SLO) Reports.....                                    | 86   |
| 12. Programs Included in Analysis.....  | 87   |
| 13. Means and Standard Deviations for Assessments.....  | 94   |
| 14. Alignment of Student Learning Outcomes with Unit Key Assessments.....                                 | 95   |
| 15. Milestone Measures from CEHD Rubrics.....   | 105  |

|   |     |
|---|-----|
| 16. Outcome, Level 1, and Level 2 Variables (Hierarchical Linear Modeling).....                         | 108 |
| 17. Unconditional Models to Measure Variability.....  | 108 |
| 18. One-Way Random Effects ANOVA for Outcomes.....  | 110 |
| 19. Intraclass Correlation Coefficients (ICCs).....   | 114 |
| 20. Contextual Models.....  | 115 |
| 21. Percentage of Between Program Variance Explained in Contextual Models.....                          | 116 |
| 22. CEHD Conceptual Framework Aligned with Candidate Knowledge, Skills, and<br>Dispositions (2007)..... | 118 |
| 23. Model Constructs, Measured Items, and Item Description.....   | 119 |
| 24. Model Fit Statistics (RMSEA, TLI, CFI, and RMSEA).....  | 132 |
| 25. Model 5 & 6: AIC and BIC Comparison.....  | 132 |
| 26. Standardized Regression Weights (Loadings) for Final Model.....                                     | 134 |
| 27. Squared Multiple Correlations for Final CF Model.....   | 135 |

LIST OF FIGURES

| FIGURE   | PAGE |
|--|------|
| 1. University of Louisville, CEHD Assessment System Feedback Loop..... | 12   |
| 2. Confirmatory Factor Analysis Model.....                             | 85   |
| 3. Base Model for Confirmatory Factor Analysis (CFA).....              | 123  |
| 4. Model 2 for Confirmatory Factor Analysis (CFA).....                 | 126  |
| 5. Model 3 for Confirmatory Factor Analysis (CFA).....                 | 127  |
| 6. Model 4 for Confirmatory Factor Analysis (CFA).....                 | 128  |
| 7. Model 5 for Confirmatory Factor Analysis (CFA).....                 | 130  |
| 8. Model 6 for Confirmatory Factor Analysis (CFA).....                 | 133  |
| 9. Mid Year SLO Report Connection Prompts.....                         | 160  |

## CHAPTER I

### INTRODUCTION

This study will serve as a metaevaluation of the University of Louisville's and College of Education and Human Development's (CEHD) teacher education assessment system, which was designed to support continuous improvement through data-driven decision-making, aligned with accreditation standards (NCATE, 2008; CAEP, 2013). Metaevaluation is the process of evaluating an evaluation or evaluation system to ensure quality in evaluation approaches, processes, and implementation (Scriven, 1969). Ultimately, this metaevaluation is a mechanism to identify strengths and weaknesses of the assessment system aligned with the purpose of the system. Therefore, the purpose of this metaevaluation is to drive improvement of the CEHD assessment system and to better support continuous improvement of teacher education programs at the University of Louisville.

Evaluation, by definition, varies from research because of the contextual nature in which evaluation takes place (Fitzpatrick, Sanders, & Worthen, 2011). Due to the contextual nature of the CEHD's teacher education assessment system or evaluation system, background of the institution and system itself are necessary to understand the components of the assessment system. The metaevaluation was be conducted with the same awareness of context, which limits generalizability of the findings from specific

research questions outlined for this study; however the process has the potential to inform the field of evaluation and assessment of approaches to ensuring quality within evaluation and assessment practices.

### **University of Louisville**

The University of Louisville is an urban institution situated within the city of Louisville, Kentucky. The university was established in 1798 and currently consists of the Brandeis School of Law, College of Arts and Sciences, College of Business and Public Administration, College of Education and Human Development, Graduate School, Kent School of Social Work, School of Dentistry, School of Medicine, School of Music, School of Nursing, School of Public Health and Information Sciences, and the J.B. Speed School of Engineering. In Fall 2013, University of Louisville enrollment was approximately 22,529 students, with 16,151 undergraduate students and 5,620 graduate students. The student population (Fall 2013) was 74.30% White, 10.25% Black, 4.89% Non-Resident Alien, 3.60% Asian, 3.42% Hispanic, 2.98% Two or More Races, 0.35% Unknown, 0.16% American Indian/Alaska Native, and 0.60% Native Hawaiian/Other Pacific Islander. (University of Louisville, 2013)

The mission of the university is to be a premier, nationally recognized metropolitan research university. The mission focuses on five strategic areas of education experience: research, creative and scholarly activity; accessibility, diversity, equity, and communication; partnerships and collaborations; and institutional effectiveness of programs and services. Further, the mission focuses on a commitment to the liberal arts and sciences and the development of the diverse community and citizens in each of these five areas (University of Louisville, 2013).

The University of Louisville is accredited by the Southern Association of Colleges and Schools (SACS) and the National Council for the Accreditation of Teacher Education (NCATE). The university's Quality Enhancement Plan (QEP) to address institutional quality and effectiveness, aligned with SACS expectations, is Ideas to Action (i2a): Using Critical Thinking to Foster Student Learning and Community Engagement. The university's i2a critical thinking initiative is grounded in the Paul-Elder Critical Thinking Framework which focuses on a student's ability to improve the quality of his or her thinking by actively applying elements of thought (reasoning) and intellectual standards (i.e., breadth, depth, logic, and significance) (Paul and Elder, 2001).

### **College of Education and Human Development (CEHD)**

The College of Education and Human Development is one of twelve units in the University of Louisville. The CEHD mission and vision are embedded within the vision of the institution to be a premier metropolitan research university. The CEHD offers programs in the Departments of Early Childhood and Elementary Education (ECEE); Educational & Counseling Psychology, Counseling and College Student Personnel (ECPY); Health & Sports Sciences (HSS); Leadership, Foundations and Human Resource Education (ELFH); Middle and Secondary Education (MISE); and Special Education (SPED).

The University of Louisville offers initial certification teacher education programs, as well as advanced educator preparation programs. Initial certification programs include both Bachelors of Science (B.S.) and Masters of Arts in Teaching (MAT). MAT programs serve students who have earned a bachelors degree in a specific content area and want to pursue a teaching certificate. Initial certification teacher

preparation programs are offered in four departments of ECEE, HSS, MISE, and SPED. University of Louisville students enrolled in initial teacher education programs within these units are referred to as candidates to differentiate higher education students from the B-12 students in school and community settings. Advanced educator preparations include the Master of Education (M.Ed.) in Teacher Leadership (with Specialization), Endorsement for English as a Second Language, Master of Education (M.Ed.) in School Counseling, Educational Specialist in School Leadership, Doctor of Education (Ed.D.) in School Leadership, and Master of Science (M.S.) in School Social Work.

### **CEHD Conceptual Framework.**

The CEHD's conceptual framework, which was revised in 2008, directly aligns with the university's QEP and critical thinking initiative. The conceptual framework serves as the framework for development and revision of programs, courses, curriculum, assessments, and the CEHD's assessment system. The alignment of the CEHD assessment system with the conceptual framework and, ultimately, the university's QEP ensures a streamlined approach to continuous improvement and institutional effectiveness that supports both unit and university level accountability.

The CEHD's conceptual framework includes the three constructs of Inquiry, Action, and Advocacy. As defined in the conceptual framework, "Under the construct of *Inquiry*, and through active engagement and skilled training in multiple methods of rigorous *Research*, candidates in the CEHD develop the knowledge, skills, and dispositions to become *Critical Thinkers*" (CEHD, 2008, p. 18). "Under the construct of *Action*, and through routine, continual, and pervasive *Practice* – whether this be in the areas of pedagogy and instructional leadership, counseling, or research – candidates in

the CEHD develop the knowledge, skills, and dispositions to become *Problem Solvers* in the community” (CEHD, 2008, p. 19). “Under the construct of *Advocacy* and through dedicated, committed *Service* to their peers, university, community, and world candidates in the CEHD develop the knowledge, skills, and dispositions to become *Professional Leaders*” (CEHD, 2008, p. 20). As shown in Table 1, the conceptual framework constructs, as qualities reflected in candidates, are critical thinking, problem solving, and professional leadership. Research, practice, and service represent the constructs as applied in teacher education candidates.

Table 1

*CEHD Conceptual Framework Aligned with Candidate Knowledge, Skills, and Dispositions (2007, p. 17)*

|   |  |  |  |
|---|--|--|--|
| Conceptual Framework Constructs           | Inquiry  | Action   | Advocacy   |
| Constructs as Learned and Applied         | Research   | Practice   | Service  |
| Constructs Reflected in Candidates        | Critical Thinkers  | Problem Solvers  | Professional Leaders   |
| Unit Dispositions Reflected in Candidates | Exhibits a disposition to inform practice through inquiry and reflection | Exhibits a disposition to critique and change practice through content, pedagogical, and professional knowledge. | Exhibits a disposition to affirm principles of social justice and equity and a commitment to making a positive difference. |



## **CEHD Accreditation**

The University of Louisville is accredited by the National Council for the Accreditation of Teacher Education, which includes all educator preparation programs within the College of Education and Human Development (CEHD). In addition to NCATE accreditation, initial certification and advanced educator preparation programs are accredited by the Kentucky Education Professional Standards Board (EPSB). NCATE and EPSB set guidelines and standards for teacher education programs within the state of Kentucky.

The University of Louisville underwent a site visit by a joint state and national Board of Examiners team in the Fall of 2008. At the time of the visit, NCATE had six standards that were used to evaluate universities under review. Standard 2, Assessment System and Unit Evaluation, was the standard most associated with the requirements of an assessment system. Educator preparation programs were required to have a system designed to collect and analyze data on performance of applicants, candidates in programs, and graduates. The target criteria for standard 2.c for the 2008 site visit, reflected the transition to the use of outcomes assessment for the purpose of continuous improvement.

“The unit has fully developed evaluations and continuously searches for stronger relationships in the evaluations, revising both the underlying data systems and analytic techniques as necessary. The unit not only makes changes based on the data, but also systematically studies the effects of any changes to assure that programs are strengthened without adverse consequences. Candidates and faculty

review data on their performance regularly and develop plans for improvement based on the data.” (NCATE)

NCATE recently merged with the Teacher Education Accreditation Council (TEAC) to form the Council for the Accreditation of Educator Preparation (CAEP), effective *July 2013*. CAEP accreditation continues to focus on the educator preparation units within the University of Louisville, which primarily resides in the CEHD. The University of Louisville must therefore adhere to the new CAEP standards (2013). The focus of this study will be on the standards specific to assessment systems and the utilization of assessment data in teacher education programs.

### **Assessment System**

The CEHD has a comprehensive assessment system designed to track teacher candidate performance data and support continuous program improvement. The assessment system consists of Hallmark Assessment Tasks (HATs), Unit Key Assessments, the Continuous Assessment and Records Documentation System (CARDS), and additional course and program specific assessments developed by CEHD administrators, faculty, and staff. Assessments are administered in the web-based performance assessment tool, LiveText©. Transitional data are captured in the university’s PeopleSoft© system and reported through Blackboard Analytics© (formerly iStrategy©). The system was designed in alignment with the NCATE standards and the CEHD conceptual framework.

Hallmark Assessments Tasks or HATs are the baseline of assessment across all educator preparation and human development programs in the CEHD and are mapped to standards and outcomes specific to the programs that the courses serve. In the early

development of the assessment system, faculty developed crosswalks of professional standards to their courses and assignments to define HATs and identified the standards that would be assessed in each HAT. Once each HAT was defined, faculty developed the rubrics to assess the HATs. These HATs were posted electronically in LiveText©. For each HAT, CEHD students are required to submit the assigned artifact electronically through LiveText©, and then the instructor of record completes an electronic standards-based rubric to evaluate the student's work. The use of LiveText© generates electronic assessment reports that provide data on how well CEHD students are performing on the standards and outcomes aligned with courses and programs.

CARDS (Continuous Assessment and Records Documentation System) is the system for tracking key assessments and milestone data. CARDS is designed to track teacher candidate performance across the phases of admissions, mid-program, and exit, which is referred to as CARDS 1 (admissions), CARDS 2 (mid-program), and CARDS 3 (exit) for initial certification programs (Bachelors and Masters of Arts in Teaching). Advanced educator program candidate performance data are tracked across CARDS 4-6 (Masters of Education and Endorsement Programs) and CARDS 7-9 (Doctoral Programs). Within these checkpoints or phases, performance assessment is captured on norm-referenced exams, admissions interviews, the Unit Key assessments, and other assessments as defined by the program.

There are 10 Unit Key Assessments across all educator preparation programs in the CEHD. These Key Assessments are consistent across all programs; however, the mode of assessment may vary by degree type (e.g., Bachelors versus Masters of

Education) and Concentration (e.g., Elementary Education versus Middle and Secondary Education). The CEHD’s 10 Unit Key Assessments are listed in Table 2.

Table 2

*University of Louisville, College of Education and Human Development 10 Unit Key Assessments for Educator Preparation*

| Unit Key Assessment    | Focus of Unit Key Assessment                              |
|------------------------|---|
| Unit Key Assessment 1  | Norm-referenced National Examination of Content Knowledge |
| Unit Key Assessment 2  | Evaluation of Content Knowledge                           |
| Unit Key Assessment 3  | Professional Standards                                    |
| Unit Key Assessment 4  | Evidence of Planning                                      |
| Unit Key Assessment 5  | Clinical Practice   |
| Unit Key Assessment 6  | Impact on P-12 Student Learning                           |
| Unit Key Assessment 7  | Ideas to Action Holistic Construct Rubric                 |
| Unit Key Assessment 8  | Ideas to Action Unit Dispositions Rubric                  |
| Unit Key Assessment 9  | Diversity   |
| Unit Key Assessment 10 | Technology  |

The Unit Key Assessments are embedded across multiple phases of a teacher candidate’s program and in some cases assessments are measured across all phases. A sample CARDS chart is provided in Appendix A. The assessment data are captured for all teacher candidates across the three phases and recorded in PeopleSoft© Milestones. PeopleSoft© is the University of Louisville’s data management system and offers a tool for tracking milestone completion and grades for teacher candidates. The milestones panel allows the CEHD to apply the appropriate milestones to a candidate’s academic record based on his or her program. The milestones can then be marked as completed, in progress, or not completed. For milestones requiring a score, a milestone grade can be applied. The CEHD assessment system uses a scale of 3, 2, 1 within the CARDS grading scheme with 3 representing Target, 2 representing Acceptable, and 1 representing Unacceptable or Needs Improvement.

All educator preparation programs are provided summary data for the 10 Unit Key Assessments on an annual basis and faculty systematically analyze, discuss, and report on the data. This process occurs during the same time that the university's annual SACS-aligned Student Learning Outcomes (SLOs) reporting process takes place. Each program completes an SLO report which requires programs to define their outcomes, align their outcomes with program goals, define the measures and targets for those outcomes, and then report findings from those measures to the Office of Academic Planning and Accountability. Programs are then asked to provide an action plan for the coming year. In addition to the university requirements, program faculty in the CEHD are also asked to reflect on the previous year's plan for improvement and discuss how they addressed their action plan. The data captured through Hallmark Assessment Tasks (HATs) and CARDS are used to support Student Learning Outcome (SLO) reports. Educator preparation programs align their outcomes with the 10 Unit Key Assessments. The Unit Key Assessments then serve as the student learning outcomes. For each student learning outcome, program faculty then define the instruments and measures that are used to generate data pertaining to that particular outcome and then report the findings from those measures. Findings are reported in percentages of CEHD students achieving performance levels of target, acceptable, and unacceptable. Program faculty are encouraged to use the findings to make data-based decisions in their action plans.

The intent of the assessment system is to drive continuous improvement. Since the last accreditation visit in 2008, significant changes have been documented for the artifacts and rubrics used for the 10 Unit Key Assessments at both the initial certification

and advanced educator preparation levels. This has been due to the emphasis on NCATE Standard 3.c, which emphasizes the use of the assessment system for program improvement and the need to revise the assessment to address program and CEHD student needs.

The College Educator Preparation Committee (CEPC), which consists of faculty within the educator preparation unit of the university, provides a platform for program faculty to share proposed and piloted revisions associated with the assessment system and to receive feedback. Further, the University Educator Preparation Committee (UEPC), which is a committee consisting of provost office administrators, educator preparation faculty, CEHD administrators, CEHD staff, school partners, and university partners from the College of Arts and Sciences, School of Social Work, School of Medicine, and others, has provided a mechanism to share the assessment system's structure and processes, as well as data captured through the system that pertains to content knowledge assessment results impacted by and with partners. The UEPC meetings have provided opportunities to review Praxis II (ETS©) Content Examination scores for CEHD teacher candidates. Faculty from the College of Arts and Sciences are instrumental in digesting and applying these data to support continuous improvement as CEHD candidates take content specific courses from Arts and Sciences faculty. These groups and the work of the faculty develop the core of the CEHD's feedback loop as shown in Figure 1.

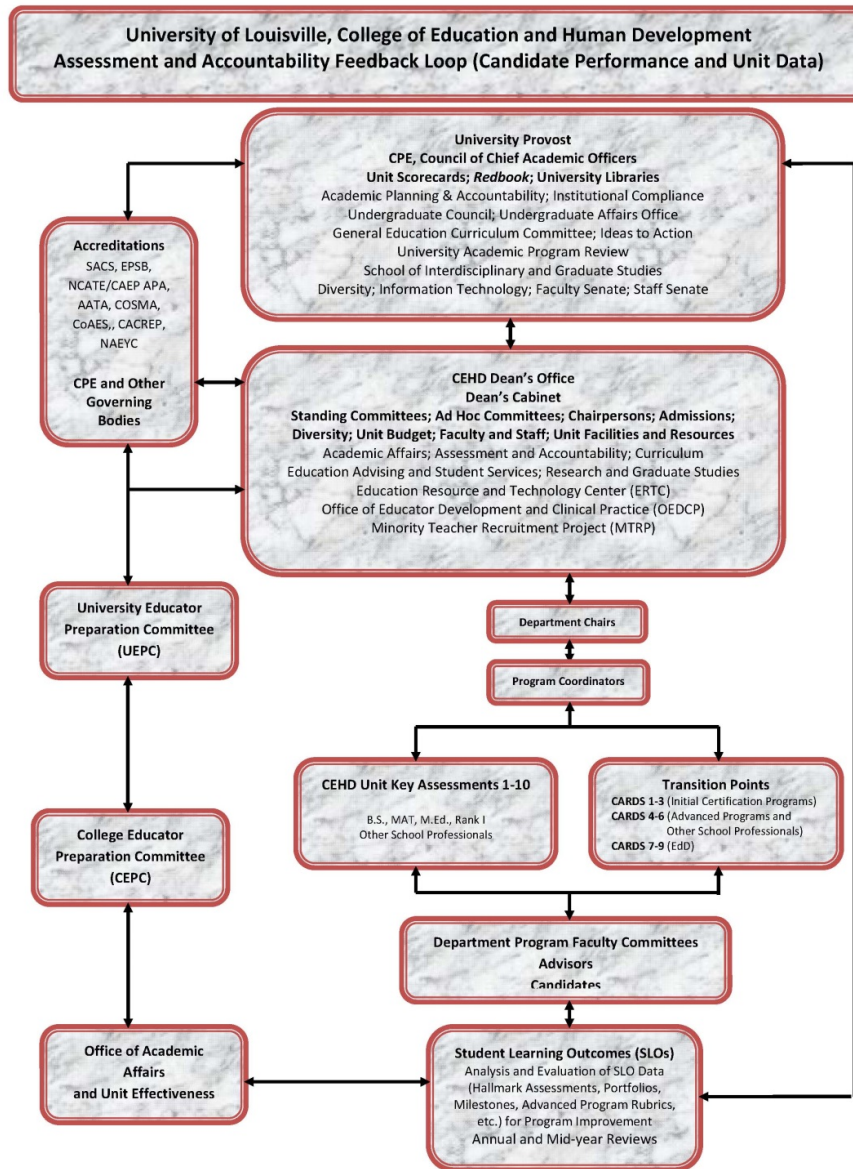


Figure 1. University of Louisville, College of Education and Human Development (CEHD) Assessment System Feedback Loop. Feedback loop provides a diagram of the flow of data and information within the CEHD's assessment system to both internal and external stakeholders.

Figure 1 highlights the process and work flow that ensure the assessment system is capturing the types of data that are needed to understand the impact and effectiveness

of programs on teacher candidates. The feedback loop demonstrates the direct intention of the assessment system to drive closing the loop through using systematic data to make decisions, inform practice, and drive program improvement and revision.

### **Purpose of Study**

This study will serve as a metaevaluation of the CEHD's assessment system. Metaevaluation is defined as evaluation of an evaluation or evaluation system (Scriven, 1969). The metaevaluation is internal and formative which Stufflebeam highlights as being a proactive guide to evaluation (2011). The metaevaluation will identify strengths and weaknesses of the system and drive improvement within the system itself (Stufflebeam, 2011; Wentling & Klit, 1973). The focus of the metaevaluation is to ensure that the CEHD assessment system is providing valid and reliable data related to CEHD student knowledge, skills, and dispositions that support faculty in ensuring continuous improvement and data driven decision-making, in alignment with the Council for Accreditation of Educator Preparation (CAEP) standards (2013) and the Joint Committee on Standards for Educational Evaluation (JCSEE) Program Evaluation Standards (2011). Further, this study seeks to contribute to the assessment and evaluation field by demonstrating one institution's process for closing the loop in the development, implementation, and revision of the assessment system with the goal of improvement.

Past NCATE standards and the new CAEP standards released in 2013 both highlight the need for continued review and revision of the assessment system (NCATE, 2008, and CAEP, 2013). As part of the process for continuous improvement and in response to the new standards, the CEHD wants to be proactive in addressing any potential weaknesses and continue to strengthen the current system and the processes



associated with it. Table 3 provides the CAEP standards that are most closely associated with the assessment system. These standards will help to drive the metaevaluation process.

Table 3

*CAEP Standards (2013)*

| Standards  |  |
|--|--|
| Standard 3: Candidate Quality, Recruitment and Selectivity | <p>3.3 Educator preparation providers establish and monitor attributes and dispositions beyond academic ability that candidates must demonstrate at admissions and during the program. The provider selects criteria, describes the measures used and evidence of the reliability and validity of those measures, and reports data that show how the academic and non-academic factors predict candidate performance in the program and effective teaching.</p> <p>3.4 The provider creates criteria for program progression and monitors candidates' advancement from admissions through completion. All candidates demonstrate the ability to teach to college and career-ready standards. Providers present multiple forms of evidence to indicate candidates' developing content knowledge, pedagogical content knowledge, pedagogical skills, and the integration of technology in all of these domains.[ii]</p>  |
| Standard 4: Program Impact                                 | <p>4.1 The provider documents, using multiple measures, that program completers contribute to an expected level of student-learning growth. Multiple measures shall include all available growth measures (including value-added measures, student-growth percentiles, and student learning and development objectives) required by the state for its teachers and available to educator preparation providers, other state-supported P-12 impact measures, and any other measures employed by the provider.</p> <p>4.2 The provider demonstrates, through structured and validated observation instruments and student surveys, that completers effectively apply the professional knowledge, skills, and dispositions that the preparation experiences were designed to achieve.</p> <p>4.3 The provider demonstrates, using measures that result in valid and reliable data and including employment milestones such as promotion and retention, that employers are satisfied with the completers' preparation for their assigned responsibilities in working with P-12 students.</p> |

Table 3 Continued

|  |   |
|--|---|
| <b>Standard 5: Provider<br/>Quality Assurance and<br/>Continuous Improvement</b> | <p>4.4 The provider demonstrates, using measures that result in valid and reliable data, that program completers perceive their preparation as relevant to the responsibilities they confront on the job, and that the preparation was effective.</p> <p>5.1 The provider’s quality assurance system is comprised of multiple measures that can monitor candidate progress, completer achievements, and provider operational effectiveness. Evidence demonstrates that the provider satisfies all CAEP standards.</p> <p>5.2 The provider’s quality assurance system relies on relevant, verifiable, representative, cumulative and actionable measures, and produces empirical evidence that interpretations of data are valid and consistent.</p> <p>5.3. The provider regularly and systematically assesses performance against its goals and relevant standards, tracks results over time, tests innovations and the effects of selection criteria on subsequent progress and completion, and uses results to improve program elements and processes.</p> <p>5.4. Measures of completer impact, including available outcome data on P-12 student growth, are summarized, externally benchmarked, analyzed, shared widely, and acted upon in decision-making related to programs, resource allocation, and future direction.</p> <p>5.5. The provider assures that appropriate stakeholders, including alumni, employers, practitioners, school and community partners, and others defined by the provider, are involved in program evaluation, improvement, and identification of models of excellence.</p> |
|--|---|

---

The new CAEP standards listed in Table 3 place heavy emphasis on validity and reliability of instruments. While this was an important component in the previous version of the NCATE standards, validity and reliability were only mentioned in standard 2 and specifically 2.3. In the new CAEP standards, validity and reliability are mentioned in standards 3.3, 4.2, 4.3, 4.4, and 5.2. The new standards also continue to demonstrate the importance of using the assessment system to evaluate candidate performance at all phases, regularly use data to make decisions regarding programs and curricula, and

ultimately develop a culture of continuous improvement. These three components help to define this study's research questions and the main focus areas of the metaevaluation.

### **Research Questions**

This study seeks to determine the extent to which the CEHD assessment system guides program improvement through the use of outcomes and standards-based data, the reliability of measures captured in the Continuous Assessment and Records Documentation System (CARDS) across educator preparation programs, and the validity of instruments used to measure critical thinking, problem solving, and professional leadership within the Unit Key Assessments for Initial Certification Teacher Preparation programs. Research questions have been defined to address each of these three areas.

To determine the extent to which the assessment system guides program improvement through the use of outcomes and standards-based data, this study identifies two research questions: (1) what types of assessments best support program improvement in teacher education programs? and (2) how are assessment data used to inform continuous improvement in teacher education? To determine the reliability of measures, this study seeks to determine (3) are the assessments used to measure teacher candidate knowledge, skills, and dispositions reliable? To determine the validity of instruments, this study seeks to determine (4) if the assessments used to measure CEHD student knowledge, skills, and dispositions are valid.

To address question 1, this study focuses on the assessments that are cited by faculty as part of their annual Student Learning Outcomes (SLO) report action plans to determine what types of assessments are informing change. Question 2 focused on the ways in which CEHD faculty apply data in making plans for continuous improvement

through their annual SLO action plans. Question 3 examined the reliability of instruments used in the CEHD assessment system and will specifically focus on the variability in assessment measures captured across initial certification programs. Question 4 examined the construct validity of measures within the CEHD assessment system that are mapped to the CEHD's conceptual framework to determine if the assessments are adequately measuring the constructs intended with the design and development of the system.

The CEHD's conceptual framework constructs are Inquiry, Action, and Advocacy (2008). The constructs as reflected in candidates' performance assessments are critical thinking, problem solving, and professional leadership. These constructs represent the latent constructs of measures in the CEHD assessment system. Faculty developed the assessment rubrics used capture CEHD student knowledge, skills, and dispositions in alignment with these three constructs.

### **Limitations**

The process of metaevaluation serves multiple purposes and has the opportunity to inform practice in teacher preparation and the evaluation field. The primary purpose of the metaevaluation is to drive continuous improvement of the CEHD assessment system to ensure that effective evaluation practices to meet program needs. Secondly, the metaevaluation can bring awareness to the process of metaevaluation for reviewing evaluations and evaluation systems to ensure that they are adequately meeting the needs of key stakeholders and are aligned with accountability expectations. In addition, this metaevaluation can serve as a model for validity and reliability testing of instruments used in outcomes assessment.

There are several limitations of this study due to the emphasis on context in evaluation and metaevaluation. The primary limitation is that the metaevaluation is focused on the effectiveness of one institution's assessment system and, therefore, is not focused on generalizability of results. Unlike traditional research, evaluation and metaevaluation focus on serving key stakeholders instead of generalizing findings to a larger population. The specificity of the assessment system to the University of Louisville and the College of Education and Human Development limits generalization of any findings.

The second major limitation is that the assessment system was designed to address specific components of the institution's conceptual framework, which is unique to this institution. While critical thinking, problem solving, and professional leadership are qualities that may be reflected in other institutions' assessment systems, the definitions of these constructs can vary and may be applied uniquely in different contexts. Use of this study to support measurement of critical thinking, problem solving, and professional leadership should ensure alignment with framework definitions.

Third, the assessments used to capture outcomes data on teacher candidate performance within the assessment system are unique to this institution. Many of the assessments are aligned with state and federal standards and educational reform documents; however, the assessments are designed explicitly for the assessment of artifacts and teacher candidate work as outlined by the program faculty in the CEHD at the University of Louisville. The rubrics were all developed by CEHD faculty and were not designed for use outside of the college.

Fourth, the evaluation was conducted by an internal evaluator, which introduces potential bias because of prior knowledge and involvement in the assessment system. To address potential bias, the evaluation was closely aligned with the Council for Accreditation of Educator Preparation (CAEP) Standards, Stufflebeam's structure of metaevaluation (2001), and Stufflebeam's Metaevaluation Checklist (1999).

To further address concerns regarding an internal evaluator, members of the dissertation committee, members of the CEHD's Office of Academic Affairs and Unit Effectiveness team, and the Executive Director of the University of Louisville's Office of Institutional Effectiveness reviewed this study or components of this study to ensure accuracy, transparency, and credibility of the process. The Executive Director of the university's Office of Institutional Effectiveness further served as a member check to address subjectivity of this study. These individuals serve as experts in assessment, evaluation, higher education accountability, institutional research, research methodology, teacher education accreditation and accountability, and teacher education program curricula. The expertise and the roles of these individuals will be further addressed in Chapter 3 as part of the metaevaluation process.

### **Significance of Study**

Metaevaluation serves two important functions: (a) to determine effectiveness of evaluation and evaluation systems, and (b) to review the role of evaluation in general (Scriven, 1969). Despite the limitations of this study and the contextual nature of this metaevaluation, the study has the potential to inform future evaluation and metaevaluation practices. Evaluation serves as a mechanism to address accountability, so therefore it is essential to review the assessment system with the purpose of

understanding the impact of evaluation, as well as determine best practices to support quality evaluation processes.

For the University of Louisville, College of Education and Human Development, this metaevaluation will help to support faculty, staff, administrators, and school district partners in revising the assessment system to ensure that the system aligns with program goals and expectations, and is in alignment with public, government, and accreditation related accountability. The review will take a systematic approach to identifying strengths and weaknesses of the system to inform future work and support decision-making in the CEHD. This study demonstrates the devotion of the CEHD to continuous improvement and a commitment to high quality programs and services that support the preparation of future educators.

The requirements for assessment systems are consistent across all NCATE /CAEP accredited institutions. This metaevaluation can potentially inform other colleges and universities with the same or similar accountability requirements about not only the importance of evaluation, but the importance of designing, implementing, and reviewing evaluation approaches, procedures, and impacts. In alignment with leaders in the evaluation field, such as Michael Scriven and Daniel Stufflebeam, metaevaluation is an essential component of any evaluation and helps to build credibility of the evaluation itself through critical and strategic review. It is in this spirit that this study was conceived and undertaken.

## **Definitions**

The primary terms used in this study are defined as follows:

1. **Accountability** – Accountability is the responsibility of higher education institutions to key constituents for the development, production, and utilization of resources (Hubbell, 2007).

2. **Action** – Action is the application of knowledge through practice (CEHD’s Conceptual Framework, 2007).

3. **Advocacy** – Advocacy is associated with the understanding of the impact of inequity and developing a professional philosophy that reflects making a positive difference for all students (CEHD’s Conceptual Framework, 2007)

4. **Assessment System** – Assessment system pertains to the systematic assessment of teacher candidate knowledge, skills, and dispositions, review of data from those assessments, and faculty reflection on data to support data-based decision-making.

5. **Continuous Assessment Records and Documentation System (CARDS)** – CARDS pertains to the University of Louisville, College of Education and Human Development’s (CEHD) assessment system and specifically the transitional assessment of teacher candidate knowledge, skills, and dispositions.

6. **Continuous Improvement** – Continuous improvement pertains to a culture of evaluation in which key stakeholders of the organizations engage in the development, implementation, and revision of the evaluation to ensure quality programs, products, and services (Taylor-Ritzler et al., 2013).

7. **Critical Thinking** – Critical thinking is “that mode of thinking – about any subject, content, or problem – in which the thinker improves the quality of his or her



thinking by skillfully taking charge of the structures inherent in thinking and imposing intellectual standards upon them” (Paul-Elder 2008).

8. Evaluation – Evaluation is the process of assessing a program or product to judge merit or worth or define strengths and weaknesses to improve quality (Fitzpatrick, Sanders, & Worthen, 2011).

9. Inquiry – Inquiry is the exploration, invention, and discovery of knowledge (Bibens, 1980).

10. Metaevaluation – Metaevaluation is the evaluation of an evaluation or evaluation system to support quality in evaluation practices (Scriven, 1969; Stufflebeam, 2011).

11. Problem Solving – Problem solving is the application, testing, and sharing of acquired knowledge in multiple settings (Shulman, 2006).

12. Professional Leadership – Professional leadership is the act of applying knowledge and skills, with a disposition towards social justice and equality and a commitment to making a positive difference (CEHD’s Conceptual Framework, 2007).

13. Teacher candidate – Teacher candidates are higher education students enrolled in programs leading to Birth-12 teaching certificate.

## CHAPTER II

### REVIEW OF LITERATURE

#### **Accountability in Higher Education**

Accountability in higher education pertains to the responsibility of institutions to key constituents for the development, production, and utilization of resources (Hubbell, 2007). Evaluation is a means to address accountability of programs and services in alignment with standards, policies, research, and market as defined by key stakeholders or constituents. Accountability is a significant driver for evaluation and of the approach, process, and implementation of evaluation in higher education (Madaus & Stufflebeam, 1984). Accountability to key stakeholders and engagement in evaluation helps to ensure quality of programs and services provided by institutions (Mero, Guidice, & Anna, 2006). In higher education, accountability requires institutions to regularly evaluate programs and services to address professional standards, policy reform, educational initiatives, educational research, and educational market to meet stakeholder expectations.

Higher education institutions are held accountable to a number of audiences including state and federal government, accrediting agencies, and the public. As a need to regulate quality within higher education institutions, federal and state government, as well as accrediting agencies, play an integral part in setting standards and ensuring quality of services provided by institutions. Further, the public drives accountability as a

consumer of higher education (Marchand & Stoner, 2012; Zemsky et al., 2005). Government, accreditation agencies, and the public represent three of the largest constituent or stakeholder groups for higher education accountability and drive evaluation to ensure quality of programs and services.

Accountability requirements are intended to improve quality in higher education through transparency of budgets, attainment of educational outcomes, research productivity, and institutional effectiveness. Government organizations, accrediting bodies, and the public define the expectations for these areas through policy reform, accreditation standards, accreditation reviews, and through consumer demand. Transparency of these areas increases the demands for efficiency and effectiveness in the use of resources and the programs and services that are offered (Findlow, 2008). The review of programs and services through the lens of policy reform, accreditation and consumer demand are all forms of evaluation in higher education (Fitzpatrick, Sanders, & Worthen, 2011).

While accountability and evaluation have been present in higher education since the founding of the first institutions (Thelin, 2011), many argue that there have been significant shifts in accountability and evaluation towards increased transparency in recent years (Mehta, 2013; Webber & Boehmer, 2008). Accountability has come to be viewed by some as compliance and regulation rather than evaluation to guide continuous improvement due to the prescriptive and quantitative nature of stakeholder expectations (Chouinard, 2013; Eaton, 2012; and Sibolski, 2012). Yet others argue that the shifts towards increased transparency and prescriptive guidelines are essential to driving evaluation and quality control within all organizations (Carmen, 2013).

Key stakeholders and constituents have held higher education institutions accountable since the founding of the first institutions (Thelin, 2011). Early development of external boards, significant financial contributors, and religious organizations represent a few of the primary stakeholders that have influenced accountability throughout time (Thelin, 2011). Funding was one of the original sources of accountability throughout history, forcing institutions of higher education to be responsible to stakeholders that provided support through financial allocations or donation of land and facilities. These stakeholders had the authority to make judgments about quality of programs and services, as well as define standards for quality aligned with their expectations for higher education because of the reliance upon their contributions to institutions. Early evidence of the connection between funding and government policy reform was the Morrill Act of 1862 and the development of land-grant colleges (Thelin, 2011). While government was not directly in the business of building or explicitly defining expectations of colleges, it was instrumental in expanding higher education throughout the United States to support a growing demand for advanced educational opportunities (Thelin, 2011).

Accreditation has been another consistent mechanism for accountability throughout higher education history (Thelin, 2011). Instead of the federal government directly regulating and judging quality in higher education, accreditation bodies were tasked with the role of reviewing higher education programs (U.S. Department of Education). In a fight against “diploma mills” during the GI Bill (1944-1956) era, the federal government agreed to an accountability system of voluntary accreditation associations (Thelin, 2011, p. 264-265). Due to concerns about relaxed standards of

accreditation and educational quality in the late 1970s there was an increase in state government involvement in accountability (Thelin, 2011, p. 340). The history of higher education accountability has been through many changes in stakeholder expectations and seen significant policy reform since the founding of the first institutions.

A major shift in accountability began in the 1970s and 1980s that led to increased use of learning outcomes and standards-based academic outcomes for measurement of quality (Mehta, 2013; Thelin, 2011). Evidence of this shift is demonstrated by the *A Nation at Risk* report by the National Commission on Excellence in Education (1983), which is referenced as a turning point in educational policy reform and accountability (Mehta, 2013). This movement continued with the No Child Left Behind (2001) reform (Mehta, 2013). The shift focused on schools being held responsible for academic outcomes, on increasing standards, increasing international competitiveness, and greater accountability at the state level (Mehta, 2013). This trend began in P-12 education and has expanded to postsecondary education, especially teacher education programs because of the direct connection to preparation of teachers and other P-12 school personnel.

Higher education institutions that provide teacher preparation programs have seen a significant push for connecting the performance of teacher education students to the performance of their students as demonstrated by the Council for Accreditation of Educator Preparation (CAEP) Standards (2013). While there are expectations for the use of data to drive continuous improvement across multiple CAEP standards, CAEP standard 4.1 explicitly requests the use of P-12 student learning outcomes data to measure impact of teacher preparation programs on P-12 students. This standard demonstrates the

increased standards in higher education accountability beyond classroom-based outcomes to workplace outcomes.

The U.S. Department of Education has pushed for increased connectivity of education outcomes to workforce placement and workforce performance data to measure quality of higher education (Mehta, 2013; U.S. Department of Education). This has further been enforced and emphasized by accreditation agencies as demonstrated by the CAEP standards. It has been argued that reform in this area requires more prescriptive and quantitative measures of evaluation, which inhibits internal stakeholder involvement in internal accountability and evaluation processes that utilize missed methods with contextual emphasis (Chouinard, 2013). Others assert that these are necessary guidelines and standards that support accountability efforts across all types of institutions (Carmen, 2013). As institutions of higher education seek to address this accountability requirement and those from other key constituents, institutions of higher education have the challenge of ensuring that they maintain a strong mission aligned with accountability requirements and that is also responsive to the educational market of stakeholders that they serve (Zemsky, Massy, & Wegner, 2005).

Market-smart and mission-centered are seen as conflicting efforts, but also require balance that institutions must have in marketing themselves to ensure revenue and continuing to meet their mission and purpose (Zemsky et al., 2005). State and federal government, as well as accrediting agencies, often drive the mission-centered components of accountability in ensuring that higher education institutions are serving the populations that are intended and serving them in a way that promotes advanced educational and professional opportunities as members of society. They drive mission-

centered components through policy reform and accreditation standards. The market component is heavily driven by public accountability and consumer demands.

While institutions are held accountable for the programs and services that they offer and ensure that they align with the vision for higher education from government and accrediting agencies, institutions are heavily impacted by the education market because of demands for resources to support programs and services. Higher education institutions have increasingly had to be concerned with funding and competition in the 21<sup>st</sup> century (Thelin, 2011). Recent decreases in the allocation of state funding to institutions of higher education have had a major impact on marketing as institutions have had to find ways to increase revenues and promote efficiency (Zemsky et al., 2005). Funding has had a direct effect on public accountability as institutions have raised tuition rates and been forced to be market-smart and mission-centered in order to attract students to their programs and identify programs that are going to capture the target audiences to ensure continued revenue. Accountability from the public, government, and accrediting agencies drives evaluation to review and regulate the quality of higher education programs to ensure a balance of market and mission.

While accountability is intended to ensure quality and drive improvement, often accountability is associated with compliance and regulation. It has been argued that accountability discourages innovation because it is seen as compliance that distracts and inhibits innovation (Chouinard, 2013; Findlow, 2008). Due to issues with buy-in and trust of accountability processes, evaluation often comes with many challenges in higher education (Bornman, 2006). Further, individual and organizational factors, such as

awareness, motivation, and competence, impact the capacity of an organization to address accountability (Taylor-Titzler et al., 2013).

Regardless of how organizations perceive accountability, it is growing and becoming more high stakes for higher education institutions. Institutions must address accountability from the public, the government, and accreditation agencies to survive in the competitive higher education market (Thelin, 2011). The following sections provide greater detail into public, government, and accreditation accountability, as they are the major forces driving evaluation in higher education.

### **Public Accountability**

Institutions of higher education are held accountable to the public, which includes current and future students, parents, employers, taxpayers, and the community as a whole. The driving forces behind public accountability include the cost associated with a college education, the demands and preparation needed for the workforce, and the continuing development of society (Marchand & Stoner, 2012). Rising costs of a college degree, student loan debt, and increased unemployment of college graduates have heightened public accountability from students and parents as consumers of higher education (Webber & Boehmer, 2008).

There is a clear relationship between cost, public interest, and accountability in higher education (Marchand & Stoner, 2012). As costs go up, public interest increases, which drives increased accountability. Consumers want to ensure that a college education will provide benefits related to employment and general quality of life (Webber & Boehmer, 2008, p. 79). Consumers have the expectation that higher education will prepare graduates to be employable after completion of their degree. For institutions of



higher education to serve these key constituents effectively, they must be responsive to consumer demands and be market smart.

Zemsky et al. (2005) explain that there is a demand for higher education by three different student consumer populations. These groups include students and parents who treat higher education as a consumable (group 1), those who are cost conscious (group 2), and those who are seeking specific skills and qualifications (group 3). Ultimately, these three groups are all part of public accountability and have many ties to both market and mission. They each have unique ways of defining and driving quality in higher education.

Students and parents who view higher education as a consumable (group 1) are interested in quality and reputation (Zemsky et al., 2005). These individuals view higher education as a financial investment in which they are looking for the best product they can get. Academic rankings, such as *U.S. News and World Report*, are one form of evaluation that addresses this area of public accountability. Beyond rankings, prestige associated with an institution heavily impacts the marketability to this group of consumers. Prestige distinguishes one university from another through quality, history, and public image (Thelin, 2011). This requires institutions to ensure that their quality of programs, services, public image, and overall effectiveness in preparing quality graduates is regularly evaluated to ensure that they are in alignment with the expectations from the public. Through this self-study, they must distinguish themselves from all other institutions.

Students who are looking to attain specific skills and qualifications (group 3) and those who are who are cost conscious (group 2) are less concerned about prestige (Zemsky et al., 2005). With these two groups, accountability drives evaluation focused on

efficiency and effectiveness of programs. Institutions must therefore provide the best quality they can and ensure that they can do it in a way that allows them to keep costs within the range of their target audience and stand out from other institutions that offer the same programs and services. This form of accountability is really all about competition as institutions try to provide the best product for the best price to meet these student needs. Evaluation in this context requires institutions to constantly review consumer educational needs, ensure quality of programs, and compare to programs offered by other institutions.

With the growth of online programs and for-profit colleges and universities in the late 20<sup>th</sup> and early 21<sup>st</sup> centuries, the competition among colleges and universities has increased public accountability focused on quality of programs, preparation for the workforce, and ultimately job placement (Thelin, 2011). It is argued that increasing demands from the public for higher education quality and accessibility are driving policy reform and accreditation reform (Hartle, 2012). Consumer awareness and consumer demand require a transparency of quality and help to define the expectations for quality (Hartle, 2012).

For the public audience seeking specific programs or courses, colleges and universities must determine what the market demands and address those specific wants and needs. This form of evaluation requires institutions to regularly examine business and industry in the surrounding areas to ensure that they are offering programs and services that align with the workforce. It has been argued that due to increased needs in the healthcare, computer, and mechanical science fields there will be an increased demand for postsecondary education graduates (Hecker, 2005). Being market-smart and catering

to this and other trends in industry and in society are both crucial components of public accountability. Higher education must be responsive to key stakeholders, which includes students, parents, and employers.

All of these student populations, regardless of their motivation in pursuing higher education drive accountability. This accountability forces institutions of higher education to evaluate their programs and services to align with their target market. Public accountability drives the review of programs and services to judge quality and alignment with the needs of the consumer, as well as business and industry. Public accountability is closely aligned with and even seen as a significant driver of accountability from accreditors and policy makers (Hartle, 2012). Public image, reputation, alignment with the workforce, and a mission focused on serving key stakeholders all significantly impact public accountability (Thelin, 2011). In order to ensure that an institution is addressing these areas of public accountability there must be time and resources devoted to evaluation directly tied to mission and market of the institution.

### **Government Accountability**

Government drives higher education accountability through educational policy reform and educational initiatives (Thelin, 2011). Government reform and education initiatives are used as a mechanism to represent the interests of the public and also to ensure accountability of funds provided to institutions through grants, tax credits, and student loans (Hartle, 2012). Although the U.S. Department of Education is not in the accreditation business, they are seen to be a regulatory force that is driving the practices associated with accreditation by increasing guidelines for education that impact accountability for higher education institutions (Eaton, 2012).

Policy reform has significantly evolved over the past thirty years to focus more on improving the quality of the U.S. educational system through advanced standards, increased outcomes assessment, and competitiveness (Mehta, 2013). As documented in the U.S. Department of Education's mission, organization overview, and policy documents, there is a heavy emphasis on increased accountability and increased transparency of outputs from universities focused around successful completion of college and job placement (U.S. DOE, 2013). Through government initiatives and policy reform, such as No Child Left Behind (U.S. DOE, 2001) and College and Career Readiness (U.S. DOE, 2010), government has defined values and expectations for states and accreditation agencies to address through their own standards and reviews of school and university programs. There is some concern that government reform is heavily impacting the traditional model of accreditation and increasing the government's role in regulation of higher education (Eaton, 2012; Hartle, 2012).

Increased accountability and policy reform at the federal level has a direct relationship with accountability at the state level. Accountability efforts at the state level are often a reflection of the federal trends and expectations (Ewell & Jones, 2006). In addition to the U.S. Department of Education, state governments have their own departments, agencies, and organizations that oversee postsecondary education institutions. There are varying approaches to accountability among states due to the nature of government structures and even state reform. Due to these varying structures, states often have different approaches to evaluating quality. The accountability climate is influenced by the mutual responsiveness between higher education institutions and state agencies (Ewell & Jones, 2006).

Government is not seen as a direct source of evaluation of higher education programs and services because they do not directly make a judgment of quality. Even though accreditation agencies and state organizations make summative judgments of quality, government accountability does have a significant role in evaluating higher education programs. Through policy reform, government agencies are identifying strengths and weaknesses to improve quality, which aligns with the definition of formative evaluation (Fitzpatrick, Sanders, & Worthen, 2011). Most importantly, they are defining standards and expectations for higher education that directly influence the standards that are used to make a judgment of quality by state agencies and accrediting review teams.

Government accountability and public accountability are very closely associated with one another (Eaton, 2012; Hartle, 2012). The rising costs of a college education, concerns of student loan debt, and potential for job placement are all factors that have led to increased transparency and accountability of higher education institutions to the government and the public (Dew, 2012; Thelin, 2011). Further, higher education institutions have also been heavily effected by cuts to federal and state budgets for education and struggling financial markets (Dew, 2012). The challenges of funding, coupled with increased competition and accountability have led to a culture of doing more with less resources. Due to this trend, higher education accountability is driving efficiency efforts in addition to striving for effectiveness (Glover & Levacic, 2007; Hubbell, 2007). Institutions must further be adaptable to increased accountability, while serving students and society with accessible and affordable higher education (Spanier, 2010).

The role of the federal government in accountability is increasingly to define the standard of quality and set guidelines for state government and accrediting agencies (Eaton, 2012). Accountability from the federal government is introduced through policy reform. State government is then more closely engaged in the evaluation of quality by applying these standards and ensuring context of policy reform within their own states (Ewell & Jones, 2006). Collectively, federal and state governments serve as driving forces for accountability and evaluation of the quality of higher education programs and services.

### **Accreditation**

Accreditation agencies are the primary mechanism for ensuring quality in higher education and have traditionally been built on the premise of defining standards, review by a team of peers, and a judgment of quality based on standards (Eaton, 2012). While government and public accountability play important roles in monitoring and regularly reviewing programs and services offered by higher education institutions, accreditation and judgment of quality is left to accrediting agencies. The U.S. Department of Education website specifically states that the Department of Education does not accredit educational institutions and/or programs; however, they do provide a list of recognized accrediting agencies. The website further states that the purpose of accrediting agencies is to ensure that education provided by higher education institutions meets an acceptable level of quality (U.S. DOE, 2013).

Accreditation has primarily been driven by the need to ensure quality; however, as a result of federal and state policy reform, accreditation agencies are now playing more of a role in compliance and regulatory affairs (Eaton, 2012; Sibolski, 2012). Eaton (2012)

outlines several factors as threats to the traditional form of accreditation due to the increased involvement of government in regulating higher education. The factors include federal investment in higher education through student grants and loans, the cost of tuition and the price of higher education, and increased public accountability. These factors have all heavily influenced the federal education agenda.

Accreditation agencies are forced to respond to government policy and education reform to defend and justify the system of accreditation in higher education (Thelin, 2011). Accreditation standards and review processes must address public accountability and government reform. The alignment of accreditation agencies with the federal education agenda is exhibited through the emphasis on outcomes assessment data and the alignment of educational outcomes with job placement and performance, which are major components of policy reform documents such as *A Nation at Risk*, *No Child Left Behind*, and *College and Career Readiness Standards* (Mehta, 2013; U.S. DOE, 2013). Government reform tied to educational outcomes and educational quality force accreditation agencies to respond with changes to accreditation standards and evaluation practices that address these components.

Accreditation agencies have more recently been pushing for outcomes-based assessment and, ultimately, towards the use of data for decision-making. Colleges and universities are required to demonstrate systematic collection and application of outcomes assessment data to support continuous improvement of programs and services (Wilkins, Young, & Sterner, 2009). Emphasis on faculty involvement and a culture of continuous assessment to drive program improvement through the use of data have continued to grow (Payne & Miller, 2009). These expectations are emphasized through

standards and assessments used to make judgments during accreditation reviews. They further help to guide higher education institutions in addressing public and government accountability.

The CAEP standards, released in 2013, demonstrate a commitment to this culture of assessment. One shift in the CAEP standards from the previous NCATE accreditation standards is the emphasis on the use of data across all standards, instead of a separate standard to address data collection and use of data (CAEP). The new standards have caught the eye of the public as moving teacher preparation forward because they are perceived to increase selectivity of teacher candidates, expand demands for the incorporation of assessment data in judging teacher performance, and, ultimately, follow graduates into the workforce to track performance and judge preparation of program quality (Ginsberg & Levine, 2013).

Accreditation agencies drive both formative and summative evaluation. Formative evaluation is used to identify strengths and weaknesses to improve quality, while summative evaluation is used to make a judgment of quality (Fitzpatrick, Sanders, and Worthen, 2010). Reviews conducted by accreditation agencies are an example of summative evaluation, as review boards use standards to make a judgment of quality of programs and services (Fitzpatrick, Sanders, and Worthen, 2010). At the same time, accreditation standards and processes require institutions to engage in continuous review of their own programs aligned with the accreditation standards to identify strengths and weaknesses. Identification of strengths and weaknesses then informs decision-making to support program improvement.



Accreditation has historically played an important role in judging and guiding quality of higher education programs and services (Thelin, 2012). That role continues to evolve as public and government accountability continue to increase (Eaton, 2012; Hartle, 2012). Ultimately, accreditation standards and review processes serve as a mechanism to address the expectations of the key stakeholders and constituents of higher education. Public, government, and accreditation related accountability requirements collectively drive evaluation and evaluation processes in higher education (Madaus & Stufflebeam, 1984).

### **Evaluation**

Evaluation is an integral part of accountability in higher education because it serves as a mechanism to review educational quality. Evaluation is defined as the process of assessing a program or product to judge merit or worth or define strengths and weaknesses to improve quality (Fitzpatrick, Sanders, & Worthen, 2011). Evaluation in higher education has been demonstrated through forms of evaluation such as outcomes assessment and standardized testing (Madaus, & Stufflebeam, 1984). Scriven (1969) describes evaluation as one of the most important functions of science because it serves as a mechanism to hold individuals and organizations accountable as well as ensure values and purpose.

Accountability drives organizations to engage in quality control and impacts consciousness of evaluation practices, procedures, and implementation (Mero, Guidice, & Anna, 2006). Accountability demands higher education institutions review alignment with expectations as defined by key stakeholders. Through strategic evaluation practices, organizations can ensure continuous improvement towards meeting internal and external

expectations. Evaluation is needed to determine quality and impact of educational programs, products, and services (Bhatt & Koedel, 2012). Evaluation should be a process of on-going assessment and monitoring of educational programs and services, with a focus on sound and successful practices to support both internal and external accountability (Madaus & Stufflebeam, 1984).

While accountability drives evaluation efforts, there are conflicting views on the impact of accountability on evaluation effectiveness among evaluators. It has been argued that accountability restricts the freedom to align evaluation with organizational context and engage valuable personnel due to preferences towards technocratic methods of measuring quality that focus more on regulation and compliance (Chouinard, 2013). The alternative perspective is that accountability is a necessary support and guide to evaluation as institutions may fail to have the knowledge and resources to effectively engage in evaluation (Carmen, 2013). The commonality between these arguments is that engagement in evaluation is necessary to drive improvement, whether it is a result of the standards and requirements put forth by external accountability or due to an organization's internal initiative.

### **Evaluation Approaches**

There are many approaches to evaluation. Some of the most commonly used approaches to evaluation include expertise-oriented, consumer-oriented, program-oriented, decision-oriented, and participant-oriented (Fitzpatrick, Sanders, and Worthen, 2011). When engaging in evaluation it is important to determine which approach best aligns with the evaluation needs. It has also been argued that the use of multiple approaches to evaluation is more beneficial than using just one type of evaluation

(Bledsoe & Graham, 2005). Just as researchers may employ the use of mixed-methods to approach a research study, evaluators may need to use multiple approaches to best evaluate a program, organization, product, or service. As with research, evaluators need to use the best approach or approaches to address the evaluation questions.

Expertise-oriented and consumer-oriented approaches are two approaches that rely heavily on knowledge of the program or service under review from an external perspective (Fitzpatrick et al., 2011). Accreditation review is an expertise-oriented approach to evaluation because the accreditation review board that evaluates program adherence to standards consists of professionals from the field who have knowledge of the field, the standards, and the expectations for quality. Consumer-oriented approaches are more related to education market and public accountability. Consumer-oriented approaches require knowledge of the consumers and their needs (Bledsoe & Graham, 2005).

Public accountability aligns very closely with consumer-oriented approaches as students, parents, and the community as a whole make a judgment of quality and value of the educational programs and services offered by an institution as consumers (Fitzpatrick et al., 2011). As consumers of higher education, students and parents review programs, costs, and many other factors when determining the best fit. The process of deciding where to go to college is a consumer evaluation.

Program-oriented approaches to evaluation include objectives-oriented, logic models, and theory based approaches to evaluating a program, process, or product. Each of these approaches focuses on the true purpose or objective of the program or product being evaluated. These approaches use the purpose to help guide the evaluation process.

Program-oriented approaches to evaluation are useful in ensuring alignment with professional standards or alignment with institutional mission as they use the underlying purpose, theory, or objective to drive the evaluation questions, processes, and procedures. Program-oriented approaches align with the expectations for evaluation and assessment of educational outcomes as they can help in defining systems of assessment to support review of educational outcomes tied to mission and professional standards. (Fitzpatrick et al., 2011)

Decision-oriented approaches are designed to support leaders in decision-making through the use of information aligned with their purpose. Decision-oriented approaches are more aligned with internal evaluation at a university through the systematic collection of data about programs, students, and outcomes to guide decisions by leaders about the future of those programs, services, and even curriculum. The current culture of accountability in higher education focuses on the use of data to drive decision-making and program improvement. This approach to evaluation is very useful in guiding institutional review boards and institutional leaders in making decisions about programs and services offered within universities. (Fitzpatrick et al., 2011)

Participant-oriented approaches are used to engage individuals within the organization in the evaluation process (Fitzpatrick et al., 2011). Participant involvement in the evaluation can positively impact sustainability and creating a culture of evaluation. This is often referred to as building evaluation capacity within the organization and aligns with creating a cycle of continuous improvement (Taylor-Ritzler et al., 2013). Engagement of internal stakeholders in the evaluation process is instrumental to

evaluation buy-in and ultimately evaluation effectiveness (Chouinard, 2013; and Payne & Miller, 2009).

With the number of different approaches to evaluation that are available and the increasing demands of accountability in higher education, institutions face the challenge of finding effective and meaningful approaches that guide improvement, without the threats of accountability inhibiting that work. Response to evaluation and accountability can vary based upon personality, perspective, and knowledge of accountability (Choinard, 2013; Mero, Guidice, & Anna, 2006). Evaluation is often approached with hesitation because of the fear that through accountability, information and data generated through evaluation will be used to create unwanted policy change and even impact funding (Bornman, Mittag, & Danie, 2006). Participatory approaches to evaluation design have been shown to improve buy-in of internal stakeholders due to their increased knowledge of the evaluation process and the accountability that drives it (Payne & Miller, 2009). While participatory approaches best support continuous formative evaluation, which is directly tied to accountability, there is also the need to address public accountability, which may align more with consumer-oriented approaches to evaluation. The approach or approaches to evaluation must be guided by the purposes of the evaluation.

### **Evaluation in Education**

Evaluation has been used as a mechanism to address and ensure accountability throughout the history of education. Evidence of early evaluation approaches include, essay examinations in Boston Grammar Schools by Horace Mann and the Board of Education, the movement for comparative research in the late 1800s led by Joseph Rice,

and then on to the introduction of accreditation review boards and professional standards. One of the most distinctive movements in educational evaluation was the push for systematic and standardized evaluation. Ralph W. Tyler, who is referred to as the Father of Educational Evaluation, was part of the movement towards outcomes for measuring student achievement, and he is deemed the first to identify these approaches to evaluation as educational evaluation (Madaus & Stufflebeam, 1984).

One commonly used approach to outcomes assessment is standardized testing. Standardized testing is a mechanism to capture longitudinal outcomes data on students. Standardized test scores have further provided data to review the impact of instructional programs on student learning and to compare outcomes across different demographic groups. The American Evaluation Association (AEA) has expressed concern about the effectiveness of standardized testing, citing a lack of impact on school improvement (Fitzpatrick, Sanders, and Worthen, 2011). The AEA encourages multiple measures of learning with a heavy emphasis on context (Fitzpatrick, Sanders, and Worthen, 2011).

Rubrics are another commonly used method of outcomes assessment. Rubrics are used to measure the level of attainment of specific outcomes and standards. Rubrics can help to define levels of performance, such as excellent, very good, good, adequate, inadequate, and poor levels for specific outcomes (King et al., 2013). Engagement of faculty in the development of outcomes-based rubrics and the evaluation design has been argued to improve ownership and confidence in and for the evaluation and is an example of participatory evaluation (King et al. 2013; Royal, 2010; Payne & Miller, 2009).

Government accountability has continued to drive the use of outcomes assessment for evaluation of educational quality through policy reform (Mehta, 2013). Policy reform

has and continues to directly impact the standards and expectations that come from accreditation agencies. Accreditation agencies, such as the Council for the Accreditation of Educator Preparation (CAEP) and the Southern Association of Colleges and Schools (SACS), are two organizations that have emphasized the need for outcomes assessment to monitor student performance and drive continuous improvement.

Assessment systems designed to capture outcomes assessment data, support discussion and analysis of data, and drive continuous improvement are one established type of evaluation system found in educational settings. The prevalence of outcomes assessment systems for evaluation in education can be evidenced by the popularity of technology systems designed to capture and report data in relation to outcomes. There are a variety of these systems in place, and institutions have been implementing them for years to support outcomes assessment (Kirchner, 2012). Universities are now struggling to go beyond simply developing and implementing these systems for data collection and assessment to ensure long-term data collection, aggregation, and reporting to make proactive and informed decisions (Wilkins et al., 2009).

### **Formative and Summative Evaluation**

Evaluation can be classified as either formative or summative. Formative evaluation is intended to inform key stakeholders of strengths and weaknesses to improve the quality of the program or product (Fitzpatrick, Sanders, & Worthen, 2011). Formative evaluation or assessment is used regularly in educational settings through feedback to students on their progress towards meeting an intended goal or through updates of curriculum to address student needs (Scriven, 1969). Formative evaluation is recommended as a proactive mechanism to support decision-making within an

organization (Stufflebeam, 2011). Engagement in formative evaluation is a necessary part of any organization to support building a culture of continuous improvement.

Summative evaluation focuses on judging merit or worth and in some cases whether or not the program should continue (Fitzpatrick, Sanders, & Worthen, 2011). Summative evaluation can be used to determine effectiveness of individuals, programs, products, and services or to make a judgment about quality or identify areas for improvement (Stufflebeam, 2001). Accreditation reviews are one example of summative evaluation as a group of professional peers use standards to make a judgment about the quality of a program or programs offered at the institution being reviewed (Eaton, 2012). This type of evaluation addresses accountability from the professional community, government organizations, and, ultimately, the public (students, parents, alumni, and employers).

Stufflebeam (2011) argues that formative evaluation should be used more for internal decision-making purposes, and an external evaluator for accountability purposes should conduct summative evaluation. The use of external evaluators for summative evaluation provides more credibility to the evaluation process, procedures, and any decisions or judgment of quality (Fitzpatrick, Sanders, & Worthen, 2011). It is further argued that formative and summative evaluation approaches should be used collectively because formative evaluation helps to drive program improvement through development and implementation, while summative evaluation gives a final judgment of the program to determine its future (Fitzpatrick, Sanders, & Worthen, 2011). Both formative and summative evaluations serve important functions in higher education in addressing accountability. Formative evaluation supports a culture of continuous improvement in



which programs and services are regularly reviewed for strengths and weaknesses and revised based on the findings. Summative evaluation serves as an official review by key constituents to determine the adequacy of programs and services. Collectively, evaluations help to address public, government, and accreditation related accountability.

### **Internal and External Evaluation**

Identifying an evaluator or group of evaluators is another complex decision in regards to evaluation. Organizations must determine whether they will benefit most from an internal or external evaluation. Evaluation conducted by an individual or individuals within the organization, can often result in bias in the evaluation process due to already present relationships, organizational politics, individual perspectives, and knowledge about the organization and the individuals in the organization (Fitzpatrick, Sanders, & Worthen, 2011). While internal evaluators are perceived to bring bias, they also have an advantage because of their awareness of the organization, specifically organizational context, knowledge, and expertise. An internal evaluator must be aware of his/her own potential bias in the process and ensure transparency in the evaluation process (Scriven, 2010).

It has been argued that participatory and collaborative approaches are necessary in the current era of evaluation due to the diversity of organizations and context (Chouinard, 2013). Participatory evaluation ensures that individuals within the organization are engaged in the evaluation process. However, in many instances organizations may not have the capacity to initiate and engage in what may be considered quality evaluation (Sanders & Nafziger, 2011). Carman (2013) argues that accountability requirements address this concern by attempting to develop standards, create common outcomes, and

even common measures of impact to guide evaluation within organizations. Choinard (2013) sees accountability as restricting participatory evaluation and creating tensions in regards to context, politics, methods, and pedagogy.

There is also concern in the evaluation field that involvement of internal stakeholders can potentially impact rigor due to the feasibility of methods for measurement and organizational capacity and evaluators must work to define a balance between rigor, feasibility, and applicability (Braverman, 2013). This balance is necessary as demonstrated by the application of the Program Evaluation Standards developed by the Joint Committee for Standards for Educational Evaluation (JCSEE). The JCSEE Program Evaluation Standards, focused on accuracy, feasibility, propriety, and utility are set as guidelines for evaluation. It is important that evaluators select the standards that are most important to the context in which the evaluation is taking place and emphasize those standards in designing and reviewing the evaluation quality (Lynch et al., 2003).

While external evaluators are perceived to bring less bias to the evaluation process, external evaluators face the challenge of learning about the stakeholders, understanding the climate and culture, and designing evaluation approaches that will meet accountability needs and align with organization capacity (Fitzpatrick, Sanders, & Worthen, 2011). The use of an external evaluator can add credibility to the evaluation from an external perspective (Fitzpatrick, Sanders, & Worthen, 2011). External evaluators are often recommended for summative evaluation in which a judgment of quality is made to ensure integrity in the evaluation decision. External review boards are used by accreditation agencies because of their need to ensure integrity of the process. While using external evaluators can help to guard against bias, external evaluators still

bring potential bias because of their own personal experiences and beliefs (Scriven, 2010).

One of the first requirements of any evaluation is to identify, disclose, and provide a plan for protecting against any potential bias that the evaluator or evaluation team may bring to the process (Fitzpatrick, Sanders, and Worthen, 2011). Scriven (2010) argues that evaluation should be initiated with no goal in mind to ensure that sound evaluation practices define the goals and ultimately the evaluation process. While it is important to try to control for potential bias, an evaluator must also disclose any potential bias that may impact the evaluation process (Fitpatrick, Sanders, & Worthen, 2011). By disclosing the potential for bias, it creates transparency of the process. Transparency of evaluation, which has been a growing trend in educational policy reform, is an important component of accountability (Mehta, 2013). Through transparency of process and product, evaluation and program quality become more credible.

With accountability, there are a number of challenges to addressing external credibility due to the potential for bias. As individuals who are internal to the evaluation find the evaluator to be credible, external audiences can question credibility of the evaluator and potential bias (Stufflebeam, 2011). Alignment of standards, as well transparency of methods and process, can help to address these concerns with credibility (Scriven, 2010). Further, use of evaluation standards, professional standards, and evaluation checklists can help to protect against potential bias by setting guidelines for the evaluation process. The JCSEE standards and the many checklists that have been developed by professional evaluators are one approach to ensuring quality in the field of evaluation (Stufflebeam, 2001).

## **Evaluation Standards, Guiding Principles, and Checklists**

The field of evaluation began to emerge as a profession aligned with research and testing in 1973; however, evaluators have struggled with defining themselves and their role, which led to issues of quality in practice and among evaluators (Madaus & Stufflebeam, 1984). Organizations such as the American Evaluation Association (AEA) and the Joint Committee on Standards for Educational Evaluation (JCSEE) are two organizations that work to bring recognition and value to the field of evaluation. These professional organizations emphasize the importance of quality in evaluation practices. These organizations have been and continue to be influential in defining ethical and pedagogical standards to support evaluation professionals in defining evaluation approaches and processes.

The Joint Committee on Standards for Educational Evaluation (JCSEE) was formed in 1975 and the original Program Evaluation Standards were published in 1988. The JCSEE Program Evaluation Standards (Appendix B) target utility, feasibility, propriety, accuracy, and accountability of evaluation. In addition to the Program Evaluation Standards, the American Evaluation Association (AEA) developed Guiding Principles for Evaluators. The Guiding Principles include systematic inquiry, competence, integrity/honesty, respect for people, and responsibilities for general and public welfare, which are all geared to ensuring ethical practice in the education field. The Program Evaluation Standards and the Guiding Principles are designed with the intent of support for sound evaluation practice and design and accountability for work in the field of evaluation. Evaluators and researchers in the evaluation field argue for the use of standards and evaluation checklists among the evaluation community due to their

ability to guide evaluation design and implementation (Cooksy & Caracelli, 2005; Lynch et al., 2003; Stufflebeam, 2001; and Wingate, 2009).

The Program Evaluation Standards serve two main purposes. First, they serve as a mechanism to guide the design and implementation of an evaluation. Second, they can be used as a mechanism to review evaluation quality, by judging adherence to the standards as evidenced by Stufflebeam's (1999) metaevaluation checklist aligned with Program Evaluation Standards. The metaevaluation checklist serves as an instrument to guide evaluation of evaluation. The Standards and Guiding Principles have further resulted in the development of many evaluation checklists which are used to support evaluation design and then the evaluation of evaluation (Stufflebeam, 2001).

Checklists and standards are mechanisms to support the evaluation process, especially in instances when there may not be dedicated resources and skilled evaluators available to ensure appropriate and effective evaluation design (Sanders & Nafziger, 2011). Standards, guidelines, and checklists can help to protect against bias in the field of evaluation. Scriven (2010) argues that evaluators should enter an evaluation goal free and use the supporting standards and checklists to take a situational approach to evaluation so that the evaluation process determines the goals. Sanders and Nafziger (2011) argue that the use of standards and guidelines helps to identify the important aspects of an evaluation to ensure that the evaluation design is specific enough to address what really needs to be addressed (p. 47).

The Western Michigan University Evaluation Center, led by evaluators and evaluation researchers such as Daniel Stufflebeam, Michael Scriven, and Lori Wingate, developed a website to provide checklists for and from evaluators. These checklists focus

on the broader aspects of design and get into the specifics of evaluation budgeting. This site serves as a valuable resource to practicing evaluators and demonstrates the prevalence and reliance on evaluation standards and guiding principles in the field of evaluation. It also demonstrates the movement towards checklists as a mechanism to ensure quality in the field of evaluation.

### **Metaevaluation**

Metaevaluation is defined as the evaluation of an evaluation or evaluation system (Scriven, 1969). The operational definition of metaevaluation is “the process of delineating, obtaining, and applying descriptive information and judgmental information - about the utility, feasibility, propriety, and accuracy of an evaluation and its systematic nature, competent conduct, integrity/honesty, respectfulness, and social responsibility – to guide the evaluation and/or report its strengths and weaknesses” (Stufflebeam, 2001, p. 185). Metaevaluation is used to review components of an evaluation system in the areas of administrative organization, personnel, objectives, evaluation, occupational programs, resources, guidance services, and stakeholders served (Wentling & Klit, 1973).

Metaevaluation is an important part of systematic quality assurance and improvement in higher education (Bornmann, Mittag, & Danie, 2006). Table 4 provides Stufflebeam’s premises for metaevaluation, which provide both rationale for conducting metaevaluation and suggestions for approach. These premises emphasize that for evaluation to be done well, evaluation processes and procedures must be reviewed to ensure accountability of the evaluator(s) and also alignment of the evaluation with the goals for quality.

Table 4

*Eight Premises for Meta-Evaluation (Stufflebeam, 2011, p. 135-136).*

---

**Premises for Meta-Evaluation**

---

1. Evaluation is the assessment of merit; thus, meta-evaluation means assessing the merit of evaluation efforts.
  2. Evaluation serves decision making and accountability; thus, metaevaluation should provide information pro-actively to support the decisions that must be made in conducting evaluation work, and meta-evaluation should provide retroactive information to help evaluators be accountable for their past evaluation work. Another way of saying this is that meta-evaluation should be both formative and summative.
  3. Evaluations should assess goals, designs, implementation, and results; thus, meta-evaluation should assess the importance of evaluation objectives, the appropriateness of evaluation designs, the adequacy of implementation of designs, and the quality and importance of evaluation results.
  4. Evaluation should provide descriptive and judgmental information and appropriate recommendations. Likewise, meta-evaluation should describe and judge evaluation work and should recommend how the evaluations can be improved and how the findings can appropriately be used.
  5. Evaluation should serve all persons who are involved in and affected by the program being evaluated; hence, meta-evaluation should serve evaluators and all the persons who are interested in their work.
  6. Evaluation should be conducted by both insiders and outsiders; generally (but not always) insiders should conduct formative evaluation for decision-making and outsiders should conduct summative evaluation for accountability. Hence, evaluators should conduct formative meta-evaluation and they should obtain external judgments of the overall merit of their completed evaluation activities.
  7. Evaluation involves the process of delineating the questions to be addressed, obtaining the needed information, and using the information in decision-making and accountability. Hence, meta-evaluators must implement three steps. The meta-evaluators must delineate questions to be addressed. They must collect, organize, and analyze the needed information. Ultimately, they must apply the obtained information to the appropriate decision-making and accountability tasks.
  8. Evaluation must be technically adequate, useful, and cost/effective, and meta-evaluation must satisfy the same criteria.
- 

The eight premises for metaevaluation defined by Stufflebeam, align with many of the principles and standards of evaluation. As with evaluation, metaevaluation can be formative or summative, internal or external, and should be guided by purpose. One of the major differences between evaluation and metaevaluation are the primary

stakeholders. With metaevaluation, the key stakeholders will include the evaluator or evaluators involved in the design, implementation, and potentially the revision of the evaluation or evaluation system, in addition to the stakeholders that are served by the original evaluation or evaluation system.

### **Metaevaluation Approach**

Stufflebeam (2001) has further outlined procedures for conducting a metaevaluation as shown in Table 5. It includes processes for engaging with stakeholders, aligning plans with standards, reviewing available data and collecting new data as needed, and judging alignment with standards. The first step requires identification and interaction with key stakeholders. In evaluation and metaevaluation, involvement of stakeholders is a critical component in the design and implementation phases to support buy-in and utilization of information obtained from the process (Royal, 2010; Payne & Miller, 2009).

Table 5

*Stufflebeam's Structure for Identifying Alternative Metaevaluation Procedures*  
(Stufflebeam, 2001, p. 191)

---

#### **Structures for Metaevaluation**

---

1. Determine and arrange to interact with the metaevaluation's stakeholders.
2. Staff the metaevaluation team with one or more qualified evaluators.
3. Define the metaevaluation questions.
4. Agree on standards, principles, and/or criteria to judge the evaluation system or particular evaluation.
5. Develop the memorandum of agreement or contract to govern the metaevaluation.
6. Collect and review pertinent available information.
7. Collect new information as needed, including, for example, on-site interview, observations, and surveys.
8. Analyze the qualitative and quantitative information.
9. Judge the evaluation's adherence to appropriate standards, principles, and/or criteria.



Table 5 Continued

10. Convey the metaevaluation findings through reports, correspondence, oral presentations, etc.

11. As needed and feasible, help the client and other stakeholders interpret and apply the findings.

---

In higher education, institutions are striving for a culture of evaluation or culture of assessment that supports continuous improvement, which evaluators argue requires involvement of key stakeholders (Braverman, 2013; Cooksy & Caracelli, 2005; King, McKegg, Oakden, & Wehipeihana, 2013; and Payne & Miller, 2009). Identification and interaction with stakeholders serves as a mechanism to understand context and purpose within the organization. Stakeholders have varying expertise, perspective, and insight into the organization and the evaluation (Fitzpatrick, Sanders, & Worthen, 2011).

After identification and interaction with stakeholders, the metaevaluator(s) should be defined. Metaevaluation can be a subjective process (Wingate, 2009). The issues with internal and external evaluation in regards to bias and credibility also apply to metaevaluation. Therefore, careful selection of an evaluator or evaluation team with the necessary knowledge and expertise is critical. Stufflebeam (2011) outlines the eleven criteria for technical adequacy of an evaluation and warns that the appropriateness of internal versus external evaluator or metaevaluator should be based on which of the eleven criteria are of the highest priority. The criteria include internal validity, external validity, reliability, objectivity, relevance, importance, scope, credibility, timeliness, pervasiveness, and cost-effectiveness. These criteria, aligned with the purpose, should drive decisions such as the appropriateness of internal versus external evaluators in approaching the metaevaluation process.

Steps three and four of Stufflebeam's (2011) steps to metaevaluation require defining the metaevaluation questions and aligning those questions with the appropriate standards. Stufflebeam (2001) recommends that metaevaluation questions focus on the merit of the evaluation and the ability of the evaluation to meet the stakeholder's needs. Once the questions have been defined, standards for judging the evaluation or evaluation system should be selected that most appropriately align with the questions. As defined by the JCSEE Program Evaluation Standards, both internal and external reviewers are encouraged to use the accuracy, utility, feasibility, and propriety standards to examine design, procedures, information collected, outcomes, and conduct of evaluators to ensure continued improvement and accountability in regards to evaluation processes and products.

Checklists are encouraged by a number of evaluators and researchers from the evaluation field for the purpose of metaevaluation (Madaus & Stufflebeam, 1984; Sanders & Nafziger, 2001; Scriven, 2011; and Stufflebeam, 2001). The development of guidelines and checklists support clear and systematic evaluation designs and processes. The JCSEE Program Evaluation Standards have also been used for simply conducting a descriptive metaevaluation to identify strengths and weaknesses and a general sense of quality (Lynch, Greer, Larson, Cummings, Harriett, Dreyfus, & Clay, 2003). Lynch et al. (2003) were intentional about selecting standards that were applicable to the particular evaluation in question as advised by Stufflebeam's approach to metaevaluation. While this may seem simplistic, it demonstrates the values of the standards to easily guide internal review processes to promote continuous improvement.

After the metaevaluation questions and standards for judging merit have been defined, Stufflebeam encourages the evaluator or evaluation team to develop a memorandum of agreement to outline the metaevaluation process. This step helps to ensure that the key stakeholders and the evaluator have mutual understanding of the expectations for the process. It also serves as a platform for disclosing any potential bias that the evaluator may bring to the process.

Steps six through eight focus on the collection and analysis of data to answer the metaevaluation questions. Step six focuses on using any data that are already available. This could be in the form of information generated by the evaluation or reports from the evaluation. Already collected information can be insightful into whether or not the evaluation is meeting its intended goals. In addition to already collected data, the evaluator may need to collect additional data to adequately judge the effectiveness of the evaluation. Collection of additional data can help to confirm or invalidate the information that was already available. Stufflebeam (2001) recommends interviews, observations, and surveys as mechanisms to capture this additional information. Once all data have been collected, it should be analyzed using the best methods for the metaevaluation questions. Finally, once the data have been analyzed, a decision about adherence to the standards defined in step four can be made.

Once all analysis has been completed and decisions about the merit of the evaluation have been made, the evaluator can present the findings to the key stakeholders. If the metaevaluation is summative, the evaluator will generally provide information about why the evaluation is or is not at an adequate level of quality. If the metaevaluation is formative in nature, it will result in identification of strengths and

weaknesses and often even follow up with recommendations of how to apply the findings.

Metaevaluation is a necessary practice to ensure quality improvement in education and it helps to ensure that evaluation is meeting the original values, goals, and objectives that were intended (Scriven, 1969). Stufflebeam's (2001) steps for metaevaluation provide a systematic process for engaging in metaevaluation. The steps are based on context, purpose, and process, which make them easy to apply in a variety of systems.

### **Evaluation versus Metaevaluation**

Evaluation and metaevaluation have many similarities; however, they serve different purposes. Both can be formative or summative and internal or external (Stufflebeam, 2001). Metaevaluation designs can employ a number of approaches, such as participatory or decision-oriented approaches, just as evaluations use these approaches. While metaevaluation and evaluation are very similar in approach and design, they are very different in their intent and purpose.

Metaevaluation serves as a mechanism to hold evaluators accountable for evaluation quality (Fitzpatrick, Sanders, & Worthen, 2011; Stufflebeam, 2001). Metaevaluation is essentially the follow-up to the evaluation to ensure that evaluation met the needs of the stakeholders and also addresses components of accuracy, feasibility, propriety, and utility as outlined in the JCSEE standards. Metaevaluation is an important component of ensuring quality in the field of evaluation (Scriven, 1969).

## **Critical Thinking, Problem Solving, and Professional Leadership**

The University of Louisville, College of Education and Human Development's (CEHD) assessment system was designed to align with the conceptual framework and the mission of the college and university. The CEHD's conceptual framework and the definition of the constructs are aligned with the University of Louisville's Quality Enhancement Plan (QEP), Ideas to Action (i2a): Using Critical Thinking to Foster Student Learning and Community Engagement (2007). i2a is grounded in the Paul-Elder critical thinking framework (2001). The purpose of that alignment is to assess student knowledge, skills, and dispositions as they relate to mission. To adequately conduct a metaevaluation of the assessment system, the conceptual framework constructs have been outlined with support from research literature representative of the university's QEP and the CEHD's conceptual framework. An understanding of the conceptual framework constructs as they are defined by the university and through literature is essential to this study.

The University of Louisville, College of Education and Human Development's mission is focused on "advancing knowledge and understanding across disciplines and constituencies to develop educational leaders who will inform policy, improve practice, strengthen communities, and address pressing social concerns" (CEHD Conceptual Framework, 2007, p. 5). In order to effectively address this mission, the conceptual framework provides an alignment of this mission to student knowledge, skills, and dispositions through the conceptual framework constructs of inquiry, action, and advocacy. The conceptual framework constructs are not only a framework for the work of the college but also for evaluation within the college. The CEHD assessment system is

grounded in the constructs of inquiry, action, and advocacy to support evaluation in alignment with the mission and vision of the college. The constructs are assessed directly and indirectly through multiple assessments embedded within the teacher education continuous assessment system to ensure that students are being assessed on critical thinking, problem solving, and professional leadership within coursework as well as field and clinical experiences.

Table 6 provides the CEHD conceptual framework as aligned with student knowledge, skills, and dispositions (2007). The constructs of inquiry, action, and advocacy as defined by the conceptual framework will be learned and applied through research, practice, and service. The constructs will be reflected in students through critical thinking, problem solving, and professional leadership. Further students will reflect dispositions that demonstrate their knowledge, skills, and dispositions to engage in continuous learning that guides practice, aligned with a commitment to making a positive difference.

Table 6

*CEHD Conceptual Framework Aligned with Candidate Knowledge, Skills, and Dispositions (2007, p. 17)*

|                                    |                   |                 |                      |
|------------------------------------|-------------------|-----------------|----------------------|
| Conceptual Framework Constructs    | Inquiry           | Action          | Advocacy             |
| Constructs as Learned and Applied  | Research          | Practice        | Service              |
| Constructs Reflected in Candidates | Critical Thinkers | Problem Solvers | Professional Leaders |

Table 6 Continued

|   |  |  |  |
|---|--|--|--|
| Unit Dispositions Reflected in Candidates | Exhibits a disposition to inform practice through inquiry and reflection | Exhibits a disposition to critique and change practice through content, pedagogical, and professional knowledge. | Exhibits a disposition to affirm principles of social justice and equity and a commitment to making a positive difference. |
|---|--|--|--|

**Inquiry and Critical Thinking**

The CEHD conceptual framework construct of inquiry is aligned with the development of critical thinking skills (2007). Inquiry is defined as the exploration, invention, and discovery of knowledge (Bibens, 1980). Inquiry is considered to be an important component of becoming an effective teacher because of the importance of gaining new knowledge and skills to support instruction specific to the needs of students (Darling-Hammond & Baratz-Snowden, 2007). Further, educational initiatives and policy reform, such as the Greater Expectations report, push for student engagement and inquiry-based learning to support problem-solving and knowledge development (American Association of Colleges and Universities, 2002).

Inquiry-based learning has been shown to positively impact students’ ability to think critically (Karantzas, Avery, Macfarlane, Mussap, Tooley, Hazelwood, & Fitness, 2013). Through inquiry, teacher candidates engage in questioning and reflection to support independent development of important knowledge and skills (Fielding, Kameenui, & Gersten, 1983). Questioning is an essential component of becoming an active learner and critical thinker because it drives thinking about a specific question or topic (Elder & Paul, 2003). Reflection is also argued to be an important phase in the

inquiry process as a mechanism to compare new information to previous knowledge and context, as well as to monitor one's own learning (Shore et al., 2012).

Shulman and Shulman (2004) define reflection as the process of evaluating, reviewing, and critiquing information. It is further argued that critical reflection and analysis of information are crucial to the learning process and becoming an active learner (Shulman & Shulman, 2004). Paul and Elder (2002) define active learning as questioning, gathering and assessing information, coming to conclusions, and adopting a point of view. As students assess the information they have gathered in order to come to conclusions, they are actively reflecting on the information from their own perspective or point of view.

Kiss and Townsend (2012) argue that inquiry is more than reflection and is a cognitive process that is gained through practice and research to construct knowledge. Research as defined by Gay and Airasian (2003) is the “formal, systematic application of the scientific and disciplined inquiry approach to the study of problems (p. 3).” Research includes defining questions, collecting information, and drawing conclusions. Inquiry is an inherent component of research because of the desire to respond to some question or solve some problem. Inquiry-based learning is exhibited through the act of research as is defined by the CEHD Conceptual Framework (2007). As teacher candidates inquire to learn, through research, they build upon their critical thinking skills and develop a disposition to inform practice through critical thinking.

Assessments related to inquiry are intended to measure a student's ability to formulate questions, gather information, and ultimately, think critically about that information (CEHD Conceptual Framework, 2007). The mapping of the CEHD's



conceptual framework constructs (2007) to the assessments in the CEHD assessment system provide a mechanism for evaluation of the CEHD mission and vision for students to become critical thinkers who inform practice through knowledge gained from inquiry-based learning and research. Assessments tied to student content, pedagogical, and professional knowledge are all a direct reflection of their learning through inquiry-based learning and research.

### **Action and Problem-Solving**

Action is the second of the three conceptual framework constructs and is aligned with problem solving ability in the CEHD's conceptual framework (2007). The action construct focuses on applying knowledge through practice. Shulman's (2006) scholarship was instrumental in shaping the CEHD conceptual framework with his focus on applying knowledge from and for multiple contexts. Thus, this construct emphasizes the importance of context and a student's ability to appropriately use the information they have gained.

Problem-solving skills build upon a student's ability to think critically and engage in active learning because of the need to first inquire about a specific question or topic and build content, pedagogical, and professional knowledge. There are many frameworks and strategies for outlining problem-solving processes and to support the development of problem-solving abilities in teacher candidates. One example includes the Carlson and Bloom (2005) framework, which focuses on the four phases of orienting, planning, executing, and checking. The orienting phase consists of digesting the problem or question and organizing one's thoughts and knowledge around the problem. The planning and executing phases collectively result in the construction of implementation of

approach and strategy to solve the question or problem. Finally, the checking phase or reflection phases promotes the review of decisions and approaches to addressing the question or problem. An additional framework for problem-solving includes decoding, representing, processing, and implementing (Singer & Voica, 2013). These phases represent the process of developing an understanding of the question or problem, applying knowledge and information to that understanding, and applying it to the situation.

These phases all align with frameworks for teaching and preparation of teachers. Duck (2000), as cited in the CEHD conceptual framework, developed 12 principles for teacher education preparation aligned with critical thinking and problem-solving. The principles focus on collaboration, looking to the future, recognizing the differences in the way students learn, and finding and sharing learning experiences (Duck, 2000). These principles exhibit the importance of using context, student characteristics, and classroom environment when applying content and pedagogical knowledge. Understanding these principles and improving problem solving ability in these settings is directly associated with practice and learning through experience as defined by the conceptual framework. These problem-solving frameworks and principles align with expectations of teachers to apply content, pedagogical, and professional knowledge in the designing, planning, implementing, managing, and critiquing of instruction to ensure a positive learning environment for students.

### **Advocacy and Professional Leadership**

Advocacy is the third construct of the conceptual framework and is aligned with professional leadership. Professional leadership, as it is defined in the CEHD conceptual

framework (2007), places an emphasis on equity and social justice and is applied through service. Social justice as it relates to diversity in school settings impacts student learning and, ultimately, student outcomes (Williams & Greanleaf, 2012). The CEHD conceptual framework construct of advocacy is to address the importance of understanding the impact of inequity and developing a professional philosophy that reflects making a positive difference for all students by advocating for beliefs and practices that ensure equity, access, and inclusion as defined in the university's Conceptual Framework and Standard for Diversity (CEHD Conceptual Framework, 2007).

For the preparation of teachers, setting and context have significant implications for the development of teaching philosophy and professional leadership. The University of Louisville is a metropolitan university situated within a large urban school district and partnerships with surrounding rural districts. To prepare students to become professional leaders in the areas that they serve, they must understand the complexities, diversity, and inequities that are present within educational settings. Bemak and Chung (2005) discuss the importance of advocacy for students, especially in diverse urban schools and high poverty schools, for the preparation of school professionals and the need for those professionals to reflect a personal interest in equity for all students. This personal interest in equity represents advocacy.

Advocacy is reflected and gained through service and through an awareness of the diversity within communities, schools, and the lives of students (CEHD Conceptual Framework, 2007). Ockerman and Mason (2012) argue that the development of advocacy and a professional disposition towards social justice requires authentic service-based learning. Through service in schools and the community, students are actively engaged in

developing their understanding of the context and diverse lives of those individuals whom they serve.

The CEHD's conceptual framework (2007) reflects an image of a teacher or school leader who will have a personal commitment to ensuring equity and social justice. This requires professional leadership exhibited through a teaching philosophy based on the understanding of the diverse lives of learners, through application of professional leadership to support learning of all students, and collaboration with others to support a positive learning environment. A teaching philosophy is a reflection of the values, beliefs, and dispositions that drive practice. Amobi (2003) argues that the development of a teaching philosophy is a continuous process that requires construction of knowledge and reflection on action to evoke perspective.

### **Critical Thinking, Problem-Solving, and Professional Leadership**

#### **Framework**

Shulman and Shulman (2004) provide a visual of how Critical Thinking, Problem Solving, and Professional Development all work together. Their model focuses on the importance of individual reflection, with reflection on understanding, practice, motivation, and vision, which are all related to one another (Shulman & Shulman, 2004). This is very much the way that teacher candidates' critical thinking skills, problem solving ability, and professional leadership are all related to one another. Critical thinking drives students to research and learn. Problem solving allows students to take into consideration all that they have learned and the experiences that they have had to apply their knowledge in the appropriate way. Professional Leadership is supported by the first two constructs as students learn more about teaching content, pedagogy, and the

stakeholders they serve they can become advocates for equity and social justice. This advocacy then helps to feed that desire to learn more and gain more experience in their field to become more effective teachers. Critical thinking skills, problem-solving ability, and professional leadership are all qualities and performance outcomes exhibited through programs, field experiences, and clinical practice in teacher education.

## CHAPTER III

### METHODS

This metaevaluation follows Stufflebeam's structure for metaevaluation as shown in Table 5 (2001). Stufflebeam's recommended framework supports a systematic approach to reviewing the assessment system. The structure outlined by Stufflebeam strongly emphasizes the importance of context for evaluation and review of evaluation. The metaevaluation first identifies stakeholders, organizational structures, and professional standards and expectations to ensure that the metaevaluation serves the intended purpose. This step-by-step approach further guards against potential bias by aligning the metaevaluation with both evaluation and professional standards that can be used as a baseline for measuring quality of the evaluation.

#### **Stakeholders**

The first task outlined by Stufflebeam is to identify and interact with the key stakeholders. The primary stakeholders for the College of Education and Human Development (CEHD) assessment system include CEHD faculty, as well as administrators and staff in the Office of Academic Affairs and Unit Effectiveness (AAUE), in the CEHD. Additional stakeholders include the students of CEHD programs, alumni of CEHD programs, partners internal and external to the university, and ultimately the P-12 students whom CEHD graduates serve. Internal and external partners

to the university often have a very direct role in providing feedback to the support development and revision of the assessment system. Teacher candidates, alumni, P-12 students, and the greater community are impacted by the quality of education that the CEHD provides.

Faculty in the CEHD are primary stakeholders in this enterprise because of their ownership and involvement in the development, implementation, and revision of the assessment system. Faculty regularly interact with the system by applying assessment instruments to evaluate CEHD student knowledge, skills, and dispositions, analyzing data to reflect on student learning, and developing plans for improvement based on the findings from data. As they are the primary users of the assessment system, the metaevaluation must reflect how the assessment system guides faculty's work.

The AAUE staff is also heavily involved in the development, implementation, revision, and oversight of the assessment system. The AAUE includes the Vice Dean of the CEHD, the Assessment Coordinator, the Accountability Coordinator, the Curriculum Coordinator, the Director of Regulatory Affairs, and the Administrative Associate. Collectively these individuals work together to support the assessment system, align the work of the CEHD with accountability requirements, and provide support structures and processes for faculty to engage in curriculum, assessment, and accountability efforts. The Vice Dean leads the work of the AAUE as an administrator with expertise in both teacher education programs and accreditation processes and procedures. Her role in the development and implementation of the assessment system is to provide vision, as well as oversight of operations and compliance in alignment with state and national expectations, policy reform, and university mission. The Assessment Coordinator manages the

technical systems used for capturing and monitoring data within the assessment system. In this role, she works closely with faculty, staff, and students to support the use of the system and data from the system. Further, she provides support and guidance to faculty in the revision of assessments, processes, and procedures that ensure systematic and meaningful assessment. The Accountability Coordinator develops and provides structures and reporting templates to engage faculty in the work of addressing accreditation and accountability requirements. Those support mechanisms are essential to ensuring systematic and strategic involvement of faculty in the work. The Curriculum Coordinator supports the faculty through CEHD and departmental curriculum committees to provide systematic and streamlined processes and procedures for addressing curriculum development, curriculum changes, curriculum gaps, and improving the quality of CEHD program curriculum to address teacher candidate needs in alignment with assessment and accountability efforts. The Director of Regulatory Affairs works closely with the university's Office of Institutional Planning and Accountability and oversees the reporting of institutional data to external stakeholders. The Administrative Associate is critical to all of these functions and supports the work of the AAUE team to ensure a cohesive team that supports the mission of the college and the unit, in alignment with expectations from external accrediting agencies.

The AAUE staff support faculty in the implementation of Unit Key Assessments in LiveText© to ensure electronic assessment of CEHD student work, aligned with professional standards. In addition, the data from the 10 Unit Key Assessments and other assessments captured through alternative systems, such as testing agencies, are recorded in the university's PeopleSoft© system. These assessments are recorded as milestones



tied to each teacher candidate's academic plan. The AAUE further supports faculty in the use of these data by providing annual reports of the Unit Key Assessments, in addition to semester level course assessments. The data are provided to faculty to support the Student Learning Outcomes (SLOs) reports that are completed annually by all program faculty within the university. The CEHD assessment system has integrated components of the assessment system into the university's processes to ensure systematic, streamlined, and sustainable practices.

The SLOs require faculty to reflect on the findings from the Unit Key Assessment data sets and other course level assessments to define an action plan for the coming year. The action plans are designed to address program needs identified from the review of data. Actions resulting from the use of data are referred to as the feedback loop, which demonstrates that the assessment system is supporting data-driven decision making to improve the quality of programs and support quality preparation of teachers. The effectiveness of the system heavily impacts the ability of the faculty to engage in the feedback loop and develop a cycle of continuous program improvement.

In addition to faculty and the AAUE, university administrators are significant stakeholders of the assessment system. NCATE and CAEP accreditations are both university accreditations. Although the majority of the teacher education programs reside in the CEHD, there are important relationships with other units in the university (College of Arts and Sciences, Kent School of Social Work, School of Medicine, and School of Music) that support content knowledge development and support the development of teacher education students. From an administrative perspective, accreditation has a significant impact on allocation of funding, public image, and marketability of programs.

## **Metaevaluator**

The second step of Stufflebeam's metaevaluation process is to identify or staff the metaevaluation team with one or more qualified individuals. The metaevaluation of the CEHD assessment system was conducted by the assessment coordinator who engages directly with the assessment system. While an internal reviewer has to be aware of potential biases because of prior involvement in the system and the college, this metaevaluation benefited from the internal review because of the significant knowledge of the system to support a detailed formative assessment that can guide the work of the college in addressing accreditation standards. While internal evaluation increases the potential for bias in the evaluation process, it also has benefits such as knowledge of the stakeholders, the culture, and the history behind the organization and the evaluation system (Fitzpatrick, Sanders, & Worthen, 2011). With a background in the field of evaluation and a strong desire to improve the quality of the assessment system to support key stakeholders, the assessment coordinator has made every effort to provide transparency in evaluation efforts aligned with CAEP standards and Stufflebeam's Metaevaluation Checklist (1999).

The assessment coordinator, author of the present study, has worked for the CEHD for over eight years. She was involved in the development, implementation, and continued evolution of the system. She is knowledgeable of data collection processes, technology functions, presentation of data to faculty, and faculty interaction with the system and with the data from the system. As an internal evaluator, she provided benefit to the study from having historical knowledge and expertise in this area. During her

tenure as assessment coordinator, she has pursued the Doctor of Philosophy degree in Educational Leadership and Organizational Development with an emphasis on Evaluation. Her background in the evaluation field, ensured alignment of this study with sound evaluation practices. Further, the assessment coordinator aligned the metaevaluation with Stufflebeam's (2001) structure for metaevaluation, CAEP standards (2013), and the Joint Committee on Standards for Educational Evaluation (JCSEE) Program Evaluation Standards (2011).

During the metaevaluation process, the Assessment Coordinator worked closely with the AAUE team, as well as her dissertation committee members, and the Executive Director of the university's Office of Institutional Effectiveness. These individuals helped to ensure accuracy, transparency, and credibility of the metaevaluation process. Feedback from each of these key constituents was incorporated into the discussion of findings and further applied in the application and next steps as a result of this study.

At completion of the study, a review of relevant CAEP standards (2013) and the JCSEE Program Evaluation Standards (2011) was incorporated into discussion to ensure that the metaevaluation did not stray from the true purpose or intent, which is to improve the quality of the assessment system. This ensured alignment of the study with state and national expectations for teacher education program assessment practices and educational evaluation practices.

### **Metaevaluation Questions**

Step 3 of Stufflebeam's metaevaluation process requires the defining of the metaevaluation questions. Four research questions were selected in alignment with the CAEP standards to address the most immediate needs of the CEHD. Table 3 provides the

CAEP standards most closely aligned with the assessment system. To address these standards, there are four research questions (Table 7).

Table 7

*Research Questions*

- 
- (1) What types of assessments best support program improvement in teacher education programs?
  - (2) How is assessment data used to inform continuous improvement in teacher education programs?
  - (3) Are the instruments that are used to measure teacher candidate knowledge, skills, and dispositions in teacher education programs reliable?
  - (4) Are the instruments that are used to measure teacher candidate knowledge, skills, and dispositions valid?
- 

Questions 1 and 2 are focused on the use of the assessment system by faculty.

These questions focused on how data from the assessment system support program improvement by driving program actions. Question 1 specifically focused on the assessments faculty cite in the SLO action plans to support data-based decision-making to drive continuous improvement. Question 2 specifically looked at the ways that faculty apply data in making plans for continuous improvement in their annual SLO action plans. These questions helped to identify strengths and weaknesses in data sources and the procedures and processes for faculty engagement in analyzing and reviewing data. These questions were intended to determine how well the assessment system is meeting the needs of faculty, as key stakeholders of the evaluation system.

Questions 3 and 4 focused on the Unit Key Assessments captured through the CARDS phases. The metaevaluation only focused on assessments that are specific to CEHD programs and that are captured within the rows of rubrics designed by CEHD faculty and staff. The measures that were reviewed are listed in Table 8.

Table 8

*Milestone Measures from CEHD Rubrics*

| <b>Measure</b>                                 | <b>Description</b>   |
|--|--|
| Kentucky Teacher Standards 1                   | Rubric Row: Demonstrates Applied Content Knowledge   |
| Kentucky Teacher Standards 2                   | Rubric Row: Designs and Plans Instruction  |
| Kentucky Teacher Standards 3                   | Rubric Row: Creates and Maintains Learning Climate   |
| Kentucky Teacher Standards 4                   | Rubric Row: Implements and Manages Instruction   |
| Kentucky Teacher Standards 5                   | Rubric Row: Assesses and Communicates Learning Results   |
| Kentucky Teacher Standards 6                   | Rubric Row: Demonstrates the Implementation of Technology  |
| Kentucky Teacher Standards 7                   | Reflects on and Evaluates Teaching and Learning  |
| Kentucky Teacher Standards 8                   | Rubric Row: Collaborates with Colleagues/Parents/Others  |
| Kentucky Teacher Standards 9                   | Rubric Row: Evaluates Teaching and Implements Professional Development   |
| Kentucky Teacher Standards 10                  | Rubric Row: Provides Leadership within School/Community/Profession   |
| University of Louisville Diversity Standard 11 | Rubric Row: Understands the Complex Lives of Students and Adults in Schools and Society  |
| Student Teaching Observation (Cycle 1)         | Rubric Row: Holistic assessment of first observed lesson in student teaching   |
| Student Teaching Observation (Cycle 2)         | Rubric Row: Holistic assessment of second observed lesson in student teaching  |
| Student Teaching Observation (Cycle 3)         | Rubric Row: Holistic assessment of third observed lesson in student teaching   |
| Student Teaching Observation (Cycle 4)         | Rubric Row: Holistic assessment of fourth observed lesson in student teaching  |
| Impact on P-12 Student Learning                | Rubric Row: Reflection and analysis of student learning to demonstrate understanding of impact on p-12 student learning.   |
| Inquiry  | Rubric Row: Inquiry "... through active engagement and skilled training in multiple methods of rigorous Research candidates in the CEHD develop the knowledge, skills, and dispositions to become Critical Thinkers/" (CF, p.18) |

Table 8 Continued

|                      |  |
|----------------------|--|
| Action               | Rubric Row: Action "... through routine, continual, and pervasive Practice—whether this be in the areas of pedagogy and instructional leadership, counseling, or research—candidates in the CEHD develop the knowledge, skills, and dispositions to become Problem Solvers in the community” (CF, p. 19) |
| Advocacy             | Rubric Row: Advocacy “... through dedicated, committed Service to their community, and world candidates in the CEHD develop the knowledge, skills, and dispositions to become Professional Leaders” (CF, p. 20)  |
| Inquiry Disposition  | Rubric Row: Exhibits a disposition to inform practice through inquiry and reflection   |
| Action Disposition   | Rubric Row: Exhibits a disposition to critique and change practice through content, pedagogical, and professional knowledge  |
| Advocacy Disposition | Rubric Row: Exhibits a disposition to affirm principles of social justice and equity and a commitment to making a positive difference  |

Question 4 focused on the construct validity of assessments. All of the assessments in the CEHD assessment system were developed with alignment to the CEHD’s conceptual framework. Assessment rubric rows have been strategically mapped to the conceptual framework constructs of Inquiry, Action, and Advocacy. This review helped to determine if those measures are accurately measuring critical thinking, problem solving, and professional leadership. Table 9 provides the alignment of measures in the assessment system to critical thinking, problem solving, and professional leadership.

Table 9

*Model Constructs, Measured Items, and Item Description*

| Construct                      | Items                             | Item Description   |
|--------------------------------|-----------------------------------|--|
| <b>Critical Thinking</b>       | Inquiry                           | Rubric Row: Inquiry "... through active engagement and skilled training in multiple methods of rigorous Research candidates in the CEHD develop the knowledge, skills, and dispositions to become Critical Thinkers/" (CF, p.18)   |
|                                | Inquiry Disposition               | Rubric Row: Exhibits a disposition to inform practice through inquiry and reflection   |
|                                | Content Knowledge                 | Rubric Row: Kentucky Teaching Standard 1: Demonstrates Applied Content Knowledge   |
|                                | Reflection Evaluation             | Rubric Row: Kentucky Teaching Standard 7: Reflects On and Evaluates Teaching and Learning<br>Rubric Row: Kentucky Teaching Standard 9: Evaluates Teaching and Implements Professional Development  |
| <b>Problem Solving Ability</b> | Action                            | Rubric Row: Action "... through routine, continual, and pervasive Practice—whether this be in the areas of pedagogy and instructional leadership, counseling, or research—candidates in the CEHD develop the knowledge, skills, and dispositions to become Problem Solvers in the community" (CF, p. 19) |
|                                | Action Disposition                | Rubric Row: Exhibits a disposition to critique and change practice through content, pedagogical, and professional knowledge  |
|                                | Planning                          | Rubric Row: Kentucky Teaching Standard 2: Designs and Plans Instruction  |
|                                | Classroom Management              | Rubric Row: Kentucky Teaching Standard 3: Creates and Maintains Learning Climate   |
|                                | Implements Instruction Technology | Rubric Row: Kentucky Teaching Standard 4: Implements and Manages Instruction<br>Rubric Row: Kentucky Teaching Standard 6: Demonstrates the Implementation of Technology  |
| <b>Professional Leadership</b> | Advocacy                          | Rubric Row: Advocacy "... through dedicated, committed Service to their community, and world candidates in the CEHD develop the knowledge, skills, and dispositions to become Professional Leaders" (CF, p. 20)  |
|                                | Advocacy Disposition              | Rubric Row: Exhibits a disposition to affirm principles of social justice and equity and a commitment to making a positive difference  |
|                                | Impact Diversity                  | Rubric Row: Impact on P-12 Student Learning<br>Rubric Row: University of Louisville Standard 11: Understands the Complex Lives of Students and Adults in Schools and Society   |

Table 9 Continued

|                     |   |
|---------------------|---|
| Assessment          | Rubric Row: Kentucky Teaching Standard 5: Assesses and Communicates Learning Results              |
| Collaboration       | Rubric Row: Kentucky Teaching Standard 8: Collaborates with Colleagues/Parents/Others             |
| Leadership Standard | Rubric Row: Kentucky Teaching Standard 10: Provides leadership within school/community/profession |

**Alignment to Metaevaluation Standards**

Stufflebeam developed the Metaevaluation Checklist (1999) to support metaevaluation processes. The checklist aligns with the standards developed by the Joint Committee on Standards for Educational Evaluation (JCSEE). The checklist was used as a guide in this metaevaluation, and the standards listed in Table 10 were the most influential in determining the strengths and weaknesses of the CEHD assessment system. Stufflebeam recommends standards under each of the four areas of utility, propriety, feasibility, and accuracy. For the purpose of this metaevaluation, the standards provided in Table 10 represent items that align with the CAEP standards and with the intent of the assessment system to support faculty use of valid and reliable data to support continuous program improvement.

Table 10

*Metaevaluation Checklist Standards (Stufflebeam, 1999)*

|           | Standard                 | Description   |
|-----------|--------------------------|---|
| Propriety | (P1) Service Orientation | Standard focused on reviewing whether or not the evaluation or evaluation system is meeting the needs of the stakeholder and identifying strengths and weaknesses to provide feedback for improvement |
| Propriety | (P4) Human Interactions  | Minimizes disruption, honors time commitments of key stakeholders, and consistently relates to all stakeholders in a professional manner  |



Table 10 Continued

|             |  |  |
|-------------|--|--|
| Accuracy    | (A5) Valid Information                                   | Encourages the use of multiple measures, as well as documentation of data collection conditions and process  |
| Accuracy    | (A6) Reliable Information                                | Encourages awareness and justification of reliability with a review of factors that influence reliability  |
| Accuracy    | (A7) Systematic Information                              | Utilizes quality control protocols, systematic checks for accuracy, multiple evaluators, verification of data entry, and controlled storage of evaluation information  |
| Accuracy    | (A10) Justified Conclusions<br>(A11) Impartial Reporting | Aligns conclusions with evaluation questions and presents findings with evidence and without bias<br>Include perspectives of multiple stakeholders and engage stakeholders in the review and use of findings |
| Feasibility | (F1) Practical Procedures                                | Minimize disruption and data burden to stakeholders, employ techniques that are realistic, and when possible embed evaluation procedures in routine activities   |
| Utility     | (U7) Evaluation Impact                                   | Involves stakeholders throughout and encourages use of findings  |

The Program Evaluation Standards (2011), developed by the Joint Committee on Standards for Educational Evaluation are provided in Appendix B. The standards focus on utility, feasibility, propriety, accuracy, and accountability. The accountability standards encourage the use of the evaluation standards and other applicable standards in guiding and evaluating the evaluations and support the use of Stufflebeam’s Metaevaluation Checklist (1999).

### **Written Agreement**

Stufflebeam next suggests a written agreement to outline the expectations for the metaevaluation. In lieu of a written agreement, two administrators were asked to serve on the dissertation committee. Vice Dean Larson represents the CEHD administration, with knowledge and expertise in the field of assessment. Associate Provost Goldstein

represents the university administration with a background in institutional research and academic affairs. This ensured that the CEHD and university leadership have representation that provided feedback and general oversight of the process as outlined in this chapter. Their roles as administrators and leaders in higher education assessment practices provided valuable perspective and feedback to the metaevaluation process.

### **Collecting Information**

The next two steps of the metaevaluation were to identify existing data and collect new information as needed. The following sections outline the methods for the evaluation of the research questions.

#### **Methods**

The first two research questions were focused on how faculty use data from the CEHD assessment system to support curriculum and program decision-making. To address these two questions, a case study analysis was conducted on the existing Student Learning Outcomes (SLOs) reports. Reports from 2012, and 2013 for all educator preparation programs were reviewed. Through this case study analysis, the metaevaluation captured the number of documented uses of data related to program improvement. Further, the case study review identified themes in the types of data used and how these are used to inform change. This case study analysis sought to identify strengths and weaknesses of the reports in regards to depth of analysis, reflection, and application of findings to support continuous improvement.

The case study analysis was a collective case study, which is the review of multiple cases to gain insight into an issue (Creswell, 2012; Stake, 1995). In this case, the review of teacher preparation programs' action plans helped to identify strengths and

weaknesses in the application of data from the assessment system. The action plans were analyzed for patterns in the discussion and narrative regarding types of assessment and application of data (Gall et al., 2006). Findings derived from the action plans are reported, with a goal to gain a deeper understanding related to the impact of the CEHD assessment in driving data-based decision-making.

For the research question pertaining to reliability, the metaevaluation collected data from the CEHD assessment system exit assessments from the 2011-2012 and 2012-2013 academic years. The data included data points for each of the assessments in Table 8. Hierarchical Linear Modeling was used to determine if there were potential program level differences on each of the assessments. This helped to determine if the measures were reliable using nested models. For this study, teacher candidates are nested in programs. The following equation represents the test of significant variability between groups on the assessment being tested. The equation was used to test for significant variability on all assessments listed in table 8.

$$\text{Level 1 Equation: "Assessments from Table 8"} = \beta_{0j} + r_{ij}$$

$$\text{Level 2 Equation: } \beta_{0j} = \gamma_{00} + u_{0j}$$

Once the results of the above equation were conducted, the Intraclass Correlation Coefficient (ICC) was calculated using the equation provided below. The ICC indicated the proportion of variance that could be attributed to program. Ideally, this proportion should be low, indicating that the measures are consistent and reliable across programs, and, ultimately, assessors as faculty are often assigned to teacher candidates within specific programs aligned with their own training and research areas.

$$\text{Intraclass Correlation Coefficient (ICC)} = \tau_{00}/(\tau_{00} + \sigma^2)$$

The final research question regarding construct validity utilized structural equation modeling. Data from the assessments outlined in Table 9 were used to determine construct validity of the CEHD conceptual framework. Critical thinking, problem solving, and professional leadership are the three constructs as demonstrated in candidates and serve as the model constructs in a confirmatory factor analysis to determine how well the measures set in place were measuring the constructs.

There are five items used to measure critical thinking ability in teacher candidates of each teacher preparation program. The items include an assessment of inquiry, the candidate's disposition related to Inquiry, candidate's demonstrated content knowledge, candidate's reflection of teaching and learning, and candidate's evaluation of teaching. The Inquiry assessment is a direct reflection of the CEHD's conceptual framework as Inquiry is the construct associated with critical thinking. The Inquiry Disposition assessment is also directly aligned with the conceptual framework, as the rubric assessment is defined by the conceptual framework description of the disposition for that construct. The content knowledge assessment is a direct reflection of a teacher candidate's ability to attain information specific to his or her professional program and be able to communicate their own understanding through artifacts and work samples. Reflection requires the candidate to question and reflect on his or her own teaching and evaluate student learning. Reflection is heavily emphasized as an important component in the literature and in the CEHD conceptual framework (Fielding et al., 1983; Shore et al., 2012; Shulman & Shulman, 2004). Further, the evaluation item is related to the teacher candidate's ability to reflect on and identify goals for teaching and professional growth

which is closely aligned with the reflection assessment. Collectively, these items align with the definition of critical thinkers in the CEHD conceptual framework.

Problem solving is measured with how well a teacher candidate demonstrates application of knowledge with the Action assessment, how well a candidate exhibits a disposition that focuses on improving education through the application of their knowledge with the Action Disposition assessment, and four Kentucky Teaching Standards (KTS). The four KTS standards include Planning Instruction, Classroom Management or Creating and Maintaining Learning Climate, Implementing Instruction, and Using Technology to Support Teaching and Learning. These standards are all associated with applying knowledge through action to support teaching.

Professional Leadership is measured through seven assessment points. They include Advocacy, Advocacy Disposition, Impact, the Leadership standard, Assessment standard, Collaboration standard, and University of Louisville Diversity standard. Advocacy is directly assessed through artifacts that demonstrate a commitment to serving the community. The Advocacy Disposition assessment focuses on a teacher candidate's ability to portray a personal disposition towards equity and making a difference as an educator. The Impact assessment is a reflection on teaching and should be directly tied to these advocacy components. The Kentucky Teaching Standard 10, *Provides Leadership within School/Community/Profession*, looks specifically at teacher education students' work related to addressing leadership in the school and community setting. The assessment standard is focused on the ability to assess the learning of all students and to be able to communicate learning results to students and parents of students at varying levels of ability. The collaboration standard is also focused on supporting learning of

students at all ability levels. The diversity standard focuses on the understanding of the complex lives of P-12 students and adults in schools and society to ensure equity and social justice. All of these assessments are tied back to being a professional leader in the school and community to promote learning of students and the greater community.

In the confirmatory factor analysis (CFA) base model, critical thinking, problem solving, and professional leadership are all correlated as the exogenous variables, which are all interrelated due to the relationships defined in the conceptual framework and previously described. The conceptual framework constructs of inquiry, action, and advocacy are identified as the marker variables as they are most closely aligned with the constructs as reflected in teacher education students. The errors for inquiry, action, and advocacy are all correlated because all three items are embedded within the same rubric. The errors for inquiry disposition, action disposition, and advocacy disposition are also all correlated since they are assessed in the same rubric.

The model was run using AMOS and will be first tested for convergent validity and discriminant validity. Convergent validity determines if the items for a factor are correlated. Items within a factor should be positive and at least moderately correlated. Next, the model was checked for discriminant validity, which ensures that items within one factor are not highly correlated with items within other factors.

The model was then checked for Heywood cases, which is indicated by negative error variance or standardized loadings greater than 1. If a Heywood case was found then the results of the model could not be interpreted. If there were no Heywood cases, then the model was checked to ensure that all items had loadings at or above .40.

Finally, model fit statistics were reviewed to determine adequate construct validity.  $\chi^2$  statistic should not be significant for adequate model fit (Kline, 2011). The Root Mean Square Error of Approximation (RMSEA) should be between 0 and .05 for good, .05-.08 for adequate model fit, and .08-.10 for marginal model fit (McCallum, Browne, and Sugawara, 1996). The Tucker Lewis Index (TLI) should be .90 to .95 for acceptable model fit and .95 and above is preferred (Bentler & Bonett, 1980). The Comparative Fit Index (CFI) should also be .90 to .95 to be considered acceptable and .95 and above is preferred (Hu & Bentler, 1999).

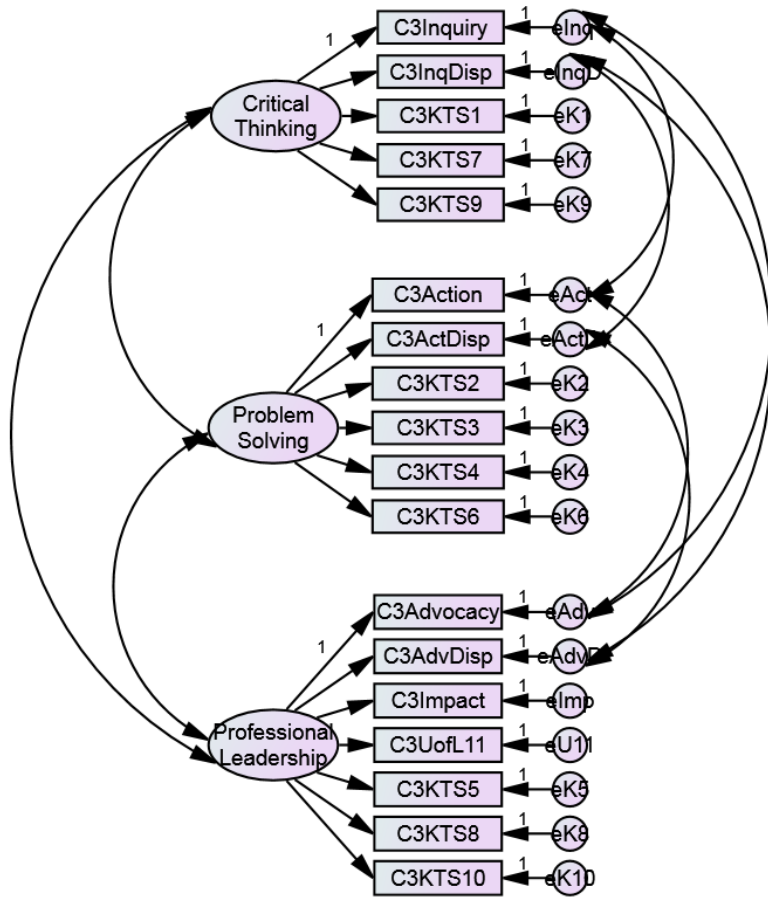


Figure 2. Confirmatory Factor Analysis Base Model. Critical Thinking, Problem Solving, and Professional Leadership are all correlated as exogenous variables, which are all interrelated due to the relationships defined in the conceptual framework. The conceptual framework constructs of Inquiry, Action, and Advocacy are all correlated because all three items are embedded within the same rubric. The errors of Inquiry Disposition, Action Disposition, and Advocacy Disposition are also correlated since they are assessed in the same rubric.



## Participants

All initial certification teacher education programs were represented in both the review of Student Learning Outcomes (SLO) reports, as well as reliability and validity testing. The SLO reports reviewed are listed in Table 11. Several of these reports represent multiple programs. Middle and Secondary Education programs complete one SLO report; however, program faculty are provided aggregate and disaggregate data for each content area to allow program faculty to review, analyze, and discuss data for action steps for continuous improvement at the program or content area level.

Table 11

### *Initial Certification Student Learning Outcomes (SLO) Reports*

| <b>SLOs (2011-2012 and 2012-2013)</b>   | <b>Program Type</b>         |
|---|-----------------------------|
| Elementary (P-5)  | Bachelors of Science        |
| Elementary dual certification with Interdisciplinary  | Bachelors of Science        |
| Early Childhood Education (IECE)  |                             |
| Elementary dual certification with Learning Behavior Disorder (LBD) or Moderate Severe Disability (MSD) | Bachelors of Science        |
| Elementary (P-5)  | Masters of Arts in Teaching |
| Health and Physical Education   | Masters of Arts in Teaching |
| Interdisciplinary Early Childhood Education (IECE)  | Masters of Arts in Teaching |
| Middle Grades   | Bachelors of Science        |
| Middle Grades   | Masters of Arts in Teaching |
| Secondary Grades  | Bachelors of Science        |
| Secondary Grades  | Masters of Arts in Teaching |
| Special Education: Learning Behavior Disorders (LBD)  | Masters of Arts in Teaching |
| Special Education: Moderate and Severe Disabilities   | Masters of Arts in Teaching |

The sample used to address the reliability and validity research questions, includes teacher candidates from the programs listed in 12. The sample included 483 teacher candidates who completed an initial certification teacher preparation program in Summer 2011 through Spring 2013. 58.8% of candidates (n=284) were enrolled in Masters of Arts in Teaching (MAT) programs. 41.2% of candidates (n=199) were enrolled in Bachelors

of Science programs. 72.54% of the MAT candidates (n=206) were enrolled in Middle and Secondary Education programs, 25% (n=71) were Elementary Education candidates, and 2.46% (n=7) were Special Education candidates. 55.78% (n=111) of the B.S. teacher education candidates were enrolled in Elementary Education programs and 44.22% (n=88) were enrolled in Middle and Secondary Education programs. The programs included in the analysis are listed in table 12. The assessment results that were used for this study were required for all teacher education students during the student teaching semester in order to successfully complete their program and therefore all teacher candidates had complete data.

Table 12

*Programs Included in Analysis*

| <b>Program</b>                               | <b>Program Type</b>         | <b>Program Concentration</b> | <b>Number of Teacher Candidates</b> |
|--|-----------------------------|------------------------------|-------------------------------------|
| Art Education                                | Masters of Arts in Teaching | Middle/Secondary             | 11                                  |
| Elementary (P-5)                             | Bachelors of Science        | Elementary                   | 32                                  |
| Elementary (P-5)                             | Masters of Arts in Teaching | Elementary                   | 71                                  |
| Elementary English                           | Bachelors of Science        | Elementary                   | 9                                   |
| Elementary French                            | Bachelors of Science        | Elementary                   | 1                                   |
| Elementary (Learning and Behavior Disorders) | Bachelors of Science        | Elementary                   | 30                                  |
| Elementary Math                              | Bachelors of Science        | Elementary                   | 16                                  |
| Elementary (Moderate and Severe Disorders)   | Bachelors of Science        | Elementary                   | 9                                   |
| Elementary Science                           | Bachelors of Science        | Elementary                   | 3                                   |
| Elementary Social Studies                    | Bachelors of Science        | Elementary                   | 11                                  |
| Health and Physical Education                | Masters of Arts in Teaching | Middle/Secondary             | 24                                  |
| Middle Grades                                | Masters of Arts in Teaching | Middle/Secondary             | 44                                  |
| Middle Grades Math                           | Bachelors of Science        | Middle/Secondary             | 12                                  |
| Middle Grades Math                           | Masters of Arts in Teaching | Middle/Secondary             | 3                                   |

Table 12 Continued

|                                     |                              |                  |    |
|-------------------------------------|------------------------------|------------------|----|
| Middle Grades Science               | Bachelors of Science         | Middle/Secondary | 1  |
| Middle and Secondary Biology        | Bachelors of Science         | Middle/Secondary | 2  |
| Middle and Secondary Chemistry      | Bachelors of Science         | Middle/Secondary | 1  |
| Middle and Secondary English        | Bachelors of Science         | Middle/Secondary | 16 |
| Middle and Secondary French         | Bachelors of Science         | Middle/Secondary | 1  |
| Middle and Secondary Spanish        | Bachelors of Science         | Middle/Secondary | 2  |
| Middle and Secondary Social Studies | Bachelors of Science         | Middle/Secondary | 16 |
| Music Education – BME               | Bachelors of Music Education | Middle/Secondary | 4  |
| Music Education – MAT               | Masters of Arts in Teaching  | Middle/Secondary | 4  |
| Music Education/Vocal or Key        | Bachelors of Music Education | Middle/Secondary | 5  |
| Music Education/BME/Instrument      | Bachelors of Music Education | Middle/Secondary | 14 |
| Secondary Biology                   | Masters of Arts in Teaching  | Middle/Secondary | 9  |
| Secondary Business                  | Masters of Arts in Teaching  | Middle/Secondary | 9  |
| Secondary Chemistry                 | Masters of Arts in Teaching  | Middle/Secondary | 8  |
| Secondary Earth Science             | Masters of Arts in Teacher   | Middle/Secondary | 2  |
| Secondary English                   | Masters of Arts in Teaching  | Middle/Secondary | 26 |
| Secondary French                    | Masters of Arts in Teaching  | Middle/Secondary | 3  |
| Secondary Math                      | Bachelors of Science         | Middle/Secondary | 15 |
| Secondary Math                      | Masters of Arts in Teaching  | Middle/Secondary | 17 |
| Secondary Physics                   | Masters of Arts in Teaching  | Middle/Secondary | 2  |
| Secondary Social Sciences           | Masters of Arts in Teaching  | Middle/Secondary | 12 |
| Secondary Social Studies            | Masters of Arts in Teaching  | Middle/Secondary | 16 |

Table 12 Continued

|                                  |                             |                   |    |
|----------------------------------|-----------------------------|-------------------|----|
| Secondary Spanish                | Masters of Arts in Teaching | Middle/Secondary  | 15 |
| Moderate and Severe Disabilities | Masters of Arts in Teaching | Special Education | 7  |

**Analyzing Findings and Judging Adherence to Standards**

This study conducted the analysis as outlined in the Methods section of Chapter 3 to answer the research questions for the metaevaluation. The study utilized mixed methods to respond to the questions. Following Stufflebeam’s (2011) steps for metaevaluation, the findings are then reported in alignment with the Program Evaluation Standards (2011) that most closely align with the questions that were determined by the research and advisor as a result of the CAEP standards (2013).

The results and discussion sections of this dissertation serve as the final report of findings, which were shared with the dissertation committee members, AAUE team members, the Executive Director of the Office of Institutional Effectiveness, and CEHD program faculty. The intent is to use the study’s findings to identify potential areas for improvement in the assessment system. The findings have the potential to inform change to support faculty engagement in the use of the assessment system to support continuous improvement of programs. The results of the study will be further shared with the College Educator Preparation Committee (CEPC) and University Educator Preparation Committee (UEPC), which serve as advisory and collaborative groups for the assessment system with input from faculty, administrators, and P-12 partners. This metaevaluation will further be submitted for public review through state, regional, and national presentations and publications.

The reliability and validity testing of instruments heavily impact next steps in the revision of the assessment system. In the event that findings require action to resolve reliability or validity concerns, faculty work groups, coordinated by the Office of Academic Affairs and Unit Effectiveness in the CEHD, will be developed to address those concerns. Faculty regularly engage in similar work groups to revise course and program assessment to support quality data collection and accreditation efforts. This study will expedite that process, by presenting the faculty with any possible weaknesses to support their engagement in the continuous improvement of the assessment system.

Further, the application of the assessment data in the Student Learning Outcomes (SLO) reports has the potential to inform change in the current AAUE templates used to guide the annual SLO reporting process. Currently, CEHD's teacher education programs have aligned their outcomes with the 10 Unit Key Assessment and ask faculty to systematically engage in analyses and discussion of those results. Findings of this study will help to identify strengths and weaknesses in that process to support potential revision of processes, reports, and templates, however, still in alignment with university guidelines for the SLO report.

Ultimately, this study serves to inform the Office of Academic Affairs and Unit Effectiveness (AAUE) staff of strengths and weaknesses in the current assessment system to best support the work of the program faculty in revising the CEHD's assessment system to better meet the evaluation needs of the CEHD. The identification of strengths and weaknesses supports the work of the AAUE staff in guiding assessment and accountability efforts across the college.

## CHAPTER IV

### RESULTS

The purpose of this metaevaluation was to identify strengths and weaknesses in the College of Education and Human Development (CEHD) assessment system. In alignment with the Council for Accreditation of Educator Preparation (CAEP) standards (2013), this metaevaluation will help to ensure that the CEHD assessment system is providing valid and reliable data collection related to teacher candidate knowledge, skills, and dispositions that support faculty in ensuring continuous improvement and data-driven decision-making. Metaevaluation is a process to ensure quality in evaluation and evaluation systems and, therefore, will guide the improvement of the CEHD assessment system to support continuous improvement of educator preparation programs.

This study addressed four research questions aligned with the CAEP standards (2013) and Joint Committee on Standards for Educational Evaluation (JCSEE), Program Evaluation Standards (2011). The four research questions include: (1) what types of assessments best support program improvement? (2) how are assessment data used to inform continuous improvement, (3) whether instruments used to measure teacher candidate knowledge, skills, and dispositions are reliable, and (4) whether the instruments used to measure teacher candidate knowledge, skills, and dispositions are valid. The first two questions are focused on identifying the assessments that faculty have cited to support program actions in their annual Student Learning Outcomes (SLO) reports and

the types of changes that faculty are citing in their action plans for programmatic annual Student Learning Outcomes (SLO) reports as a result of data collected within the CEHD's assessment system. The reliability question focused on whether there is statistically significant variability in the measurement of milestones captured across initial certification teacher education programs. The validity question focused on the assessments that have been mapped to the conceptual framework constructs of Inquiry, Action, and Advocacy in the Continuous Assessment Record and Documentation System (CARDS) milestones to determine if the assessments are adequately measuring the latent constructs of critical thinking, problem-solving, and professional leadership.

This study employed mixed methods to address the research questions. The first two questions were addressed through a qualitative case study review of the teacher preparation Student Learning Outcome (SLO) Reports from 2012 (2011-2012) and 2013 (2012-2013). Quantitative data were collected from the CEHD's assessment system to address the reliability and validity of instrumentation (research questions 3 and 4) used in the CEHD assessment system. Sampling and results of analysis are provided in following sections of this chapter.

### **Data**

The study reviewed the action plans from all initial certification Student Learning Outcome (SLOs) reports. The program SLOs analyzed for the case study analysis are included in Table 11. Middle and Secondary Education programs complete one SLO report for the Bachelor of Science (B.S.) and one for the Master of Arts (MAT) in Teaching. The program faculty are provided aggregate and disaggregate data for the Unit Key Assessments across content areas to support their discussion and analysis of data.

For the quantitative analysis to address reliability and validity of instruments, data were obtained for 483 teacher candidates who completed an initial certification teacher preparation program in the 2011-2012 or 2012-2013 academic year (Summer 2011-Spring 2013). Of the 483 teacher candidates, 58.8% of teacher candidates (n=284) were enrolled in Masters of Arts in Teaching (MAT) programs and 41.2% of teacher candidates (n=199) were enrolled in Bachelors of Science (B.S.) programs. Of the MAT candidates, 72.54% (n=206) were enrolled in Middle and Secondary Education programs, 25% (n=71) were Elementary Education teacher candidates, and 2.46% (n=7) were Special Education teacher candidates. Also, the B.S. programs consisted of 55.78% (n=111) Elementary Education teacher candidates and 44.22% (n=88) Middle and Secondary Education teacher candidates.

Demographic data were obtained for the teacher candidates included in the sample. Of the 483 teacher candidates, 80.18% were female (n=352) and 29.84% were male (n=131). Ethnicities included in the population were White (n=423, 87.58%), Black (n=29, 6.00%), Asian (n=8, 1.66%), Hispanic/Latino (n=9, 1.86%), American Indiana/Alaska Native (n=2, 0.41%), and 2.48% (n=12) were Unknown. A breakdown of the sample by program is provided in Table 12.

### **Descriptive Statistics for Quantitative Analysis**

The CEHD's 10 Unit Key Assessments are measured on a 3-point scale. The descriptive statistics for the items included in reliability and validity analysis are presented in Table 13.



Table 13

*Means and Standard Deviations for Assessments*

| <b>Variable</b>      | <b>N</b> | <b>M</b> | <b>SD</b> |
|----------------------|----------|----------|-----------|
| Inquiry              | 483      | 2.72     | .46       |
| Action               | 483      | 2.67     | .47       |
| Advocacy             | 483      | 2.64     | .48       |
| Inquiry Disposition  | 483      | 2.73     | .45       |
| Action Disposition   | 483      | 2.72     | .45       |
| Advocacy Disposition | 483      | 2.68     | .47       |
| Impact               | 483      | 2.80     | .44       |
| Cycle1               | 483      | 2.70     | .47       |
| Cycle2               | 483      | 2.78     | .42       |
| Cycle3               | 483      | 2.86     | .35       |
| Cycle4               | 483      | 2.93     | .25       |
| KTS1                 | 483      | 2.80     | .40       |
| KTS2                 | 483      | 2.75     | .44       |
| KTS3                 | 483      | 2.81     | .39       |
| KTS4                 | 483      | 2.79     | .41       |
| KTS5                 | 483      | 2.65     | .49       |
| KTS6                 | 483      | 2.67     | .47       |
| KTS7                 | 483      | 2.73     | .45       |
| KTS8                 | 483      | 2.74     | .44       |
| KTS9                 | 483      | 2.77     | .42       |
| KTS10                | 483      | 2.73     | .46       |
| UofL11               | 483      | 2.71     | .46       |

**Qualitative Case Study Analysis**

Each of the reports for the teacher preparation programs utilizes the 10 Key Assessments as the framework for their SLOS. Table 14 shows the alignment of the 10 Key Assessments to the SLOs as defined in the SLO reports for all teacher education programs. In addition to the 10 Unit Key Assessments, Hallmark Assessment Tasks (HATs) from CEHD courses serve as primary data sources to support programmatic findings about the outcomes aligned with each of the 10 Unit Key Assessments.

Table 14

*Alignment of Student Learning Outcomes with Unit Key Assessments*

| <b>Student Learning Outcome (SLO)</b>   | <b>Unit Key Assessment</b>   |
|---|--|
| Candidates demonstrate knowledge of content.  | Unit Key Assessment 1: Praxis Examination Scores (Exemplary, Pass, or Fail) and Cumulative GPA   |
| Candidate demonstrates pedagogical and professional knowledge, skills, and dispositions.                        | Unit Key Assessment 2: Kentucky Teacher Standard 1 (Demonstrates Knowledge of Content) and Professional GPA (teacher education specific course work) |
| Candidates demonstrate the ability to plan instrument.  | Unit Key Assessment 3: Kentucky Teacher Standard 2 (Designs and Plans Instruction)   |
| Candidates will demonstrate proficient performance in student teaching/clinical practice.                       | Unit Key Assessment 5: Student Teaching Observation Forms (4 Cycles)   |
| Candidates will demonstrate positive impact on P-12 student learning.   | Unit Key Assessment 6: Assessment of Impact on P-12 Student Learning   |
| Candidates will demonstrate application of the constructs of inquiry, action, and advocacy.                     | Unit Key Assessment 7: Ideas to Action Holistic Construct Rubric (Rubric rows specifically assess Inquiry, Action, and Advocacy)                     |
| Candidates will demonstrate dispositions of inquiry, action, and advocacy.                                      | Unit Key Assessment 8: Ideas to Action Unit Dispositions Rubric (Rubric rows specifically assess Inquiry, Action, and Advocacy)                      |
| Candidates demonstrate professional leadership for social justice and equity (diversity).                       | Unit Key Assessment 9: University of Louisville Standard 11 (Understands the Complex Lives of Students and Adults in Schools and Society)            |
| Candidates demonstrate ability to integrate the use of technology into their teaching and their students' work. | Unit Key Assessment 10: Kentucky Teacher Standard 6 (Demonstrates the Implementation of Technology)  |

In the annual SLO report, program faculty develop an action plan, which serves as a mechanism to apply findings from the 10 Unit Key Assessments to support continuous program improvement. As defined by the University of Louisville, Office of Academic

Planning and Accountability (OAPA), the action plan is a mechanism to “close the loop” and demonstrate application of assessment data by applying the findings or results of assessments implemented within the academic program. The SLO action plan serves as a mechanism to identify potential program improvements and to address any deficiencies identified in the findings section of the SLO report. The SLO reports further provide data on continuous improvement for SACS accreditation purposes.

This metaevaluation is specifically focused on how the CEHD’s assessment system drives continuous program improvement. The action plan, by definition, represents the mechanism for application of data and, ultimately, indicates the impact of the system. For this study, the programmatic action plans were collected from each of the initial certification teacher education programs in the CEHD. These action plans were analyzed to respond to research question 1, regarding the type of assessments that best support program improvement in teacher education and question 2, regarding the type of change resulting from assessment data from the CEHD’s assessment system.

### **Assessments Used to Support Program Improvement**

The 2012-2013 action plans for initial certification teacher education programs revealed that only one of nine action plans explicitly stated that their goals identified in the action plan were directly based upon the data from the 10 Unit Key Assessments. However, the other eight programs’ action plans provide goals related to outcomes captured in the 10 Unit Key Assessments. While the connections are not explicit, there is evidence to support that faculty are making connections between the action plan and data from the 10 Unit Key Assessments. Several programs also stated that they had the goal to improve candidate performance across outcomes, which indicates faculty are directly

referencing the 10 Unit Key Assessments associated with the SLOs. Examples of implied application of the data from the 10 Unit Key Assessments include the following:

*“Increase emphasis on KY Teacher Standard 5 (Assessment) across other courses”* (Middle and Secondary Education). – **Unit Key Assessment 5**

*“Increase explicit attention to technology as it applies to planning and teaching (i.e., how might MS or HS students use technology to explore your content are?) (KY Teacher Standard 6 (Technology) across professional courses)”* (Middle and Secondary Education). - **Unit Key Assessment 10**

*“Discuss ways to support faculty and candidates in collaborative research and inquiry – increasing a ‘research disposition’ among candidates”* (Middle and Secondary Education). – **Unit Key Assessments 7 and 8**

*“Faculty who instruct in the undergraduate program will continue to support future teacher candidates [develop content knowledge for the programs’] Praxis Content Exam”* (Health and Physical Education). – **Unit Key Assessment 1**

*“Candidates will develop a more in depth understanding of how to apply the constructs of Action, Advocacy, and Inquiry in classroom based experiences and professional dispositions”* (Special Education, Moderate and Severe Disabilities). – **Unit Key Assessments 7 and 8**

*“Candidates will develop a more in depth understanding of diversity and how to meet the diverse needs of students in their classrooms”* (Special Education, Moderate and Severe Disabilities). – **Unit Key Assessment 9**

The 2011-2012 action plans revealed similar results to the 2012-2013 action plans. While faculty are discussing the findings of the 10 Unit Key Assessments in the findings section of the SLO report, they are not explicitly citing those findings in the action plans. Sample references to the 10 Unit Key Assessments included the following:

*“Discuss ways that technology instruction impacts candidates’ performance. Specifically, discuss the following: (1) What technology should teachers and K-12 students use in schools? (2) When do we provide the necessary instruction for our candidates? and (3) How do we hold candidates accountable?”* (Middle and Secondary Education) – **Unit Key Assessment 10**

*“The faculty will continue to identify quality student teaching placement...”* (Health and Physical Education) - **Unit Key Assessment 5**

*“The faculty will continue their efforts in supporting teacher candidates in identifying appropriate technology assignments and resources they can use to support student learning”* (Health and Physical Education). – **Unit Key Assessment 10**

These action plan statements demonstrate that programs are using the 10 Unit Key Assessments to support program changes. In most instances, the relationship to the 10 Unit Key Assessments is implied and not explicitly stated by faculty. This may be a result of the design for the SLO report process or the electronic template provided to faculty. Plans for addressing this finding will be further discussed in Chapter V.

In addition to the 10 Unit Key Assessments, faculty further made references to other direct and indirect assessments of student learning. For other direct assessments, faculty made references to Hallmark Assessment Tasks (HATs) at the course level, which are used to support program improvement and also to support formative feedback to teacher candidates throughout their coursework. For indirect references, faculty referred to QMS survey results, which are student satisfaction surveys administered across all University of Louisville programs. Samples of how these data have been applied in the action plans include the following:

*“Continue to refine course HATs to ensure that diversity is integrated throughout all aspects of the program”* (Elementary Education). – **HAT Reference**

*“Continue to infuse information about preparing for professional employment opportunities (QMS survey) and job seeking skills”* (Middle and Secondary Education). – **QMS Reference**

*“Discuss ways to support faculty and candidates in addressing the “use of appropriate methods of inquiry in their field to analyze, understand, and develop effective solutions to problems (QMS survey)”* (Middle and Secondary Education). – **QMS Reference**

Collectively, these statements demonstrate that faculty use data to implement change in their programs and develop a culture of continuous improvement. While faculty are not directly referencing the assessments by name, it is implied that collectively the use of Unit Key Assessments, Hallmark Assessment Tasks (HATs), and indirect survey data (QMS) are baseline assessments that support program improvement in initial certification teacher education programs at the University of Louisville.

One other theme that arose in the action plans were references to incorporating critical thinking into courses, assignments and assessments, which align with the University of Louisville's Quality Enhancement Plan (QEP) and the CEHD's conceptual framework. This demonstrates that faculty are going beyond reflecting on their student learning outcomes' data from courses and major milestones, to making connections to the university's QEP and the CEHD's conceptual framework constructs.

In addition to the action plan, the reflection on last year's plan for improvement (action plan) included in the SLO report represents the follow-through on the items outlined in the action plans. To further support the findings that the 10 Unit Key Assessments, HATs, QMS survey data, and the university's critical thinking initiative are driving continuous improvement, the following representative statements were collected from the reflections of the previous year's action plan from the 2012 and 2013 reports.

*"The... department revised core course Hallmark Assessment Tasks rubric language to measure elements of critical thinking" (Elementary Education). – 2013*

*"Course changes were submitted to and approved by the Curriculum Committee..." (Elementary Education) – 2013*

*"Added critical thinking Elements of Reasoning to HAT rubrics; students self-assess the use of critical thinking skills in their Unit of Study" (Middle and Secondary Education) – 2013*

*“[Program] courses have been reviewed and refined by [program] faculty to ensure alignment with the [program specific] knowledge and practice”*  
(Interdisciplinary Early Childhood Education). – 2013

*“[Faculty], in response to last year’s analysis of data, have again refined the field experience and documentation to improve the connection between university learning and practical applications. Embedding additional opportunities to demonstrate understandings of formative assessment and technology integration through the field placement assignments could be a helpful step toward moving more candidates toward the target rating”* (Middle and Secondary Education). – 2012

*“The plans were completed successfully as evidenced by passing scores on the Praxis II exam by all, 100%, [degree] teacher candidates enrolled in the program”* (Health and Physical Education). - 2012

In some cases, faculty failed to reflect or report out on the implementation of the previous year’s SLO action plans. It is important to note that this is a component that was added in 2012 to further support closing the loop and promote the application of data from the CEHD’s assessment system. Further follow-up will be needed to ensure that faculty are implementing their action plans and “closing the loop” as stated in their action plans.

### **Application of Assessment Findings**

Beyond identifying which assessments are driving continuous improvement in the CEHD’s assessment system, this study focused on identifying what types of changes faculty are making as a result of engaging with assessment data. For this component of the qualitative case study, the SLO action plans were organized by the researcher into themes. The themes that were identified by this analysis include curriculum plans, faculty plans, candidate performance plans, assessment revision plans, and field/clinical experience plans.

**Curriculum plans.** Curriculum plans include references to actions such as addition of new courses, changes to course content, and modification of program requirements. These changes demonstrate that faculty are reflecting on how curriculum impacts student learning and making changes to course and program requirements to support teacher candidate development. Action plan statements that align with curriculum plans include the following:

*“Our vision beginning this year is to continue the work from the last report by ongoing program review and program revision as needed to strengthen. The [program] committee has submitted several course changes to the Curriculum Committee for approval. Our goal is to have those revisions in place for the Fall 2013 academic year”* (Interdisciplinary Early Childhood Education). – 2012

*“Provision of supervisory and administrative instruction so that graduates meet the [program specific] standards”* (Interdisciplinary Early Childhood Education). – 2012

*“Make the infusion of critical thinking explicit to candidates throughout the program”* (Middle and Secondary Education). – 2012

*“Examine and revise EDTP [course number] (Assessment Course)”* (Middle and Secondary Education). – 2013

*“Examine and revise EDTP [3 course numbers] (Technology courses) to ensure that students have opportunities to connect the technology tools they are learning to content and lesson/unit design”* (Middle and Secondary Education). – 2013

*“The faculty will work together to incorporate more diversity related assignments in the [program] curriculum. Providing a variety of field placements sites will assist with that endeavor”* (Health and Physical Education). – 2013

**Faculty plans.** Faculty plans include references to faculty development that supports student learning and faculty engagement in research. The representative statements provided below indicate that the CEHD faculty have a commitment to research both to improve their own development and to support the development of the teacher candidates within their programs. There is also a connection to ensuring that



faculty across the department collaborate to enhance their knowledge and instructional capacity.

*“New tenure-track faculty member continues to be actively engaged in grant and research efforts associated with the program and faculty members across departments”* (Elementary Education with dual certification in Interdisciplinary Early Childhood Education). – 2012

*“Enhance ELL Instruction through targeted department-wide faculty professional development”* (Interdisciplinary Early Childhood Education). – 2013

*“Discuss ways to support faculty and candidates in collaborative research and inquiry – increasing a ‘research disposition’ among candidates”* (Middle and Secondary Education). - 2013

***Candidate performance plans.*** Action plans associated with candidate performance include items such as improving content knowledge (Praxis scores) and increasing candidate performance on assessments. The items provided below align with the Student Learning Outcomes (SLO) report sections of program goals, student learning outcomes, and findings. These action plans demonstrate the connection between the data and teacher candidate learning. The action plans further support the intended purpose of the SLO report to provide faculty with an opportunity to critically look at candidate performance data, identify strengths and weaknesses, and apply those findings to support future teacher candidates’ learning.

*“Candidates will develop a more in depth understanding of how to apply the constructs of Action and Advocacy in classroom based experiences and professional dispositions”* (Special Education, Moderate and Severe Disabilities). – 2012

*“Candidates will increase performance to the Target range (75% of students at Target)”* (Special Education, Moderate and Severe Disabilities). – 2012

*“Candidates will be able to infuse various forms of technology in their planning and teaching”* (Middle and Secondary Education). - 2013

*“Candidates will know and be able to use a variety of assessment tools, both formative and summative, in planning and teaching” (Middle and Secondary Education). - 2013*

**Assessment plans.** The fourth theme that arose from the action plans was associated with the assessments used in the assessment system or supplementary assessments that are currently used or may be needed to support deeper understanding of evidence for teacher candidate learning and satisfaction. Assessment modifications align with good evaluation practices; thus, faculty are encouraged to identify needed changes to their assessments and to develop plans to systematically revise assessments. The following representative statements demonstrate that faculty are thinking critically about whether or not data are providing meaningful information about candidate learning and how assessments should be revised to provide more accurate and meaningful data related to teacher candidate knowledge, skills, and dispositions.

*“ Investigate instruments that could be used as pre and post measures of critical thinking early in Phase 3 [of field and clinical practice] and at the end of Phase 4...” (Elementary Education). - 2012*

*“Discuss ways that we collect candidate performance data and use it to impact instruction” (Middle and Secondary Education). – 2013*

*“Identify indirect measures to determine student satisfaction related to each student learning outcome (QMS data is not currently available [for this program])” (Special Education, Moderate and Severe Disabilities). – 2013*

*“Continue to refine the HAT in all courses to ensure key outcomes of the program are clearly addressed and aligned with professional standards” (Elementary Education with dual certification in Learning Behavior Disorders or Moderate and Severe Disabilities). - 2013*

**Field and clinical plans.** As demonstrated by the CAEP standards (2013), field and clinical practice are critical components to teacher preparation programs and the development of teacher candidate knowledge, skills, and dispositions. There is also heavy

emphasis on field experiences in the state of Kentucky through Education Professional Standards Board (EPSB) regulations requiring diverse field experiences and extensive field experience hours prior to student teaching. The following representative field and clinical experience items arose as the fifth and final area in which faculty described changes in their action plans and reflections on last year's plan for improvement (action plan).

*“To meet the new Kentucky KAR regulation that teacher candidates have 200 field hours prior to student teaching, hours in the school were literally calculated rather than using the approximate “1/2 day.” Our [program] candidates have more than 200 hours based on this calculation. However, it was determined that more field hours were needed for [course name] to prepare our candidates. Thus, [course name] was moved to the first semester of coursework and the 36 hours of field placement occur during their content methods semester” (Elementary Education). – 2013*

*“The faculty will continue to identify quality student teaching placement at [multiple grade levels] ...” (Health and Physical Education). – 2013*

*“There was an 8% increase (79% to 87%) of the [program] graduates who agreed or strongly agreed that the program “provided adequate opportunities for program-related fieldwork/internship/practicum experiences.” [Program] faculty have increased collaboration with the OEDCP personnel to determine areas to strengthen. [Program] candidates complete 180 hours of field component work prior to student teaching. Multiple opportunities are available for [program] teacher candidates to volunteer in community agencies and/or schools as well as attend professional development sessions (i.e. school level PD, state conferences, regional conferences, etc.)” (Health and Physical Education). - 2012*

Identification of these themes provides greater understanding of how the CEHD's assessment system is supporting data-driven decision-making. The representative statements that have been provided demonstrate that faculty are critically analyzing teacher candidate performance data to make decisions about improving curriculum, faculty development, student performance, assessments, and field and clinical experiences. Strengths and weaknesses of the above analysis are discussed in Chapter V.

## Reliability Analysis

Research question 3 was focused on the reliability of measures in the CEHD assessment system (Table 15). For the purpose of this metaevaluation, the study looked specifically at reliability across initial certification teacher education programs. As aligned with the CAEP (2013) standards, reliability is important to ensure that data from the assessment system that are used to inform decision making are reliable and can be used to identify strengths and weaknesses in student learning outcomes and to inform program improvement.

Table 15

### *Milestone Measures from CEHD Rubrics*

| <b>Measure</b>                                 | <b>Description</b>  |
|--|---|
| Kentucky Teacher Standards 1                   | Rubric Row: Demonstrates Applied Content Knowledge                                      |
| Kentucky Teacher Standards 2                   | Rubric Row: Designs and Plans Instruction   |
| Kentucky Teacher Standards 3                   | Rubric Row: Creates and Maintains Learning Climate                                      |
| Kentucky Teacher Standards 4                   | Rubric Row: Implements and Manages Instruction  |
| Kentucky Teacher Standards 5                   | Rubric Row: Assesses and Communicates Learning Results                                  |
| Kentucky Teacher Standards 6                   | Rubric Row: Demonstrates the Implementation of Technology                               |
| Kentucky Teacher Standards 7                   | Reflects on and Evaluates Teaching and Learning   |
| Kentucky Teacher Standards 8                   | Rubric Row: Collaborates with Colleagues/Parents/Others                                 |
| Kentucky Teacher Standards 9                   | Rubric Row: Evaluates Teaching and Implements Professional Development                  |
| Kentucky Teacher Standards 10                  | Rubric Row: Provides Leadership within School/Community/Profession                      |
| University of Louisville Diversity Standard 11 | Rubric Row: Understands the Complex Lives of Students and Adults in Schools and Society |

Table 15 Continued

|  |   |
|--|---|
| Student Teaching Observation (Cycle 1) | Rubric Row: Holistic assessment of first observed lesson in student teaching  |
| Student Teaching Observation (Cycle 2) | Rubric Row: Holistic assessment of second observed lesson in student teaching   |
| Student Teaching Observation (Cycle 3) | Rubric Row: Holistic assessment of third observed lesson in student teaching  |
| Student Teaching Observation (Cycle 4) | Rubric Row: Holistic assessment of fourth observed lesson in student teaching   |
| Impact on P-12 Student Learning        | Rubric Row: Reflection and analysis of student learning to demonstrate understanding of impact on p-12 student learning.  |
| Inquiry                                | Rubric Row: Inquiry "... through active engagement and skilled training in multiple methods of rigorous Research candidates in the CEHD develop the knowledge, skills, and dispositions to become Critical Thinkers/" (CF, p.18)  |
| Action                                 | Rubric Row: Action "... through routine, continual, and pervasive Practice—whether this be in the areas of pedagogy and instructional leadership, counseling, or research— candidates in the CEHD develop the knowledge, skills, and dispositions to become Problem Solvers in the community” (CF, p. 19) |
| Advocacy                               | Rubric Row: Advocacy “... through dedicated, committed Service to their community, and world candidates in the CEHD develop the knowledge, skills, and dispositions to become Professional Leaders” (CF, p. 20)   |
| Inquiry Disposition                    | Rubric Row: Exhibits a disposition to inform practice through inquiry and reflection  |
| Action Disposition                     | Rubric Row: Exhibits a disposition to critique and change practice through content, pedagogical, and professional knowledge   |
| Advocacy Disposition                   | Rubric Row: Exhibits a disposition to affirm principles of social justice and equity and a commitment to making a positive difference   |

To address this question through metaevaluation, Hierarchical Linear Modeling (HLM) was used to analyze measures captured across initial certification teacher education programs. The data sample included all teacher education program completers from the 2011-2012 and 2012-2013 academic years (Summer 2011-Spring 2013). HLM

supports analysis of data from individuals nested within groups. In this study, teacher candidates are nested in teacher education programs. To complete this analysis, each teacher candidate was assigned a program ID based upon the teacher education program in which he or she was enrolled. Table 12 provides the list of programs included in the analysis and the number of teacher candidates enrolled in each of those programs. In addition, each teacher candidate was also coded as Bachelors of Science (0) or Master of Arts in Teaching (1) and Elementary (1), Middle and Secondary (2), or Special Education (3) to support further analysis of potential variability in outcomes. As shown in Table 15, each of the measures captured on individual teacher candidates serves as the outcome measures in the HLM models and is measured on a 3-point scale (3=Target, 2=Acceptable, 1=Unacceptable). ProgramType and DegreeType serve as Level 2 variables to support further analysis of between program variability.

The rubrics used to assess each of the outcomes were developed by faculty of CEHD teacher education programs, in partnership with P-12 school partners, and field and clinical staff. Teacher candidates submit artifacts through LiveText©, and faculty complete the assigned assessment rubric electronically, at designated program points, to provide feedback to the teacher candidate on his or her performance and capture data on outcomes. The outcomes listed in Table 16 are measured and captured within individual rows of the completed rubrics.

Table 16

*Outcome, Level 1, and Level 2 Variables (Hierarchical Linear Modeling)*

| <b>Variable</b>      | <b>Level</b> | <b>Description</b>   |
|----------------------|--------------|--|
| ID                   |              | Program (1-38)   |
| Inquiry              | Outcome      | Scores coded as Target = 3, Acceptable = 2, Unacceptable – 3 |
| Action               | Outcome      | Scores coded as Target = 3, Acceptable = 2, Unacceptable – 3 |
| Advocacy             | Outcome      | Scores coded as Target = 3, Acceptable = 2, Unacceptable – 3 |
| Inquiry Disposition  | Outcome      | Scores coded as Target = 3, Acceptable = 2, Unacceptable – 3 |
| Action Disposition   | Outcome      | Scores coded as Target = 3, Acceptable = 2, Unacceptable – 3 |
| Advocacy Disposition | Outcome      | Scores coded as Target = 3, Acceptable = 2, Unacceptable – 3 |
| Impact               | Outcome      | Scores coded as Target = 3, Acceptable = 2, Unacceptable – 3 |
| Cycle1               | Outcome      | Scores coded as Target = 3, Acceptable = 2, Unacceptable – 3 |
| Cycle2               | Outcome      | Scores coded as Target = 3, Acceptable = 2, Unacceptable – 3 |
| Cycle3               | Outcome      | Scores coded as Target = 3, Acceptable = 2, Unacceptable – 3 |
| Cycle4               | Outcome      | Scores coded as Target = 3, Acceptable = 2, Unacceptable – 3 |
| KTS1                 | Outcome      | Scores coded as Target = 3, Acceptable = 2, Unacceptable – 3 |
| KTS2                 | Outcome      | Scores coded as Target = 3, Acceptable = 2, Unacceptable – 3 |
| KTS3                 | Outcome      | Scores coded as Target = 3, Acceptable = 2, Unacceptable – 3 |
| KTS4                 | Outcome      | Scores coded as Target = 3, Acceptable = 2, Unacceptable – 3 |
| KTS5                 | Outcome      | Scores coded as Target = 3, Acceptable = 2, Unacceptable – 3 |
| KTS6                 | Outcome      | Scores coded as Target = 3, Acceptable = 2, Unacceptable – 3 |
| KTS7                 | Outcome      | Scores coded as Target = 3, Acceptable = 2, Unacceptable – 3 |
| KTS8                 | Outcome      | Scores coded as Target = 3, Acceptable = 2, Unacceptable – 3 |
| KTS9                 | Outcome      | Scores coded as Target = 3, Acceptable = 2, Unacceptable – 3 |

Table 16 Continued

|             |                  |   |
|-------------|------------------|---|
| KTS10       | Outcome          | Scores coded as Target = 3, Acceptable = 2, Unacceptable – 3          |
| UofL11      | Outcome          | Scores coded as Target = 3, Acceptable = 2, Unacceptable – 3          |
| DEGREETYPE  | Level 2 Variable | Bachelors of Science (BS) = 0, Masters of Arts and Teaching (MAT) = 1 |
| PROGRAMTYPE | Level 2 Variable | Elementary = 1, Middle/Secondary = 2, Special Education = 3           |

To determine reliability, a null model or unconditional model was run for each of the outcomes provided in Table 16. The equation for each of the null models is provided in Table 17. The null model determines whether or not there is statistically significant variability in the outcome by program.

Table 17

*Unconditional Models to Measure Variability*

**Null Models**

- 
- $INQUIRY_{ij} = \gamma_{00} + u_{0j} + r_{ij}$
  - $ACTION_{ij} = \gamma_{00} + u_{0j} + r_{ij}$
  - $ADVOCACY_{ij} = \gamma_{00} + u_{0j} + r_{ij}$
  - $INQUIRYDISPOSITION_{ij} = \gamma_{00} + u_{0j} + r_{ij}$
  - $ACTIONDISPOSITION_{ij} = \gamma_{00} + u_{0j} + r_{ij}$
  - $ADVOCACYDISPOSITION_{ij} = \gamma_{00} + u_{0j} + r_{ij}$
  - $IMPACT_{ij} = \gamma_{00} + u_{0j} + r_{ij}$
  - $CYLCE1_{ij} = \gamma_{00} + u_{0j} + r_{ij}$
  - $CYLCE2_{ij} = \gamma_{00} + u_{0j} + r_{ij}$
  - $CYLCE3_{ij} = \gamma_{00} + u_{0j} + r_{ij}$
  - $CYLCE4_{ij} = \gamma_{00} + u_{0j} + r_{ij}$
  - $KTS1_{ij} = \gamma_{00} + u_{0j} + r_{ij}$
  - $KTS2_{ij} = \gamma_{00} + u_{0j} + r_{ij}$
  - $KTS3_{ij} = \gamma_{00} + u_{0j} + r_{ij}$
  - $KTS4_{ij} = \gamma_{00} + u_{0j} + r_{ij}$
  - $KTS5_{ij} = \gamma_{00} + u_{0j} + r_{ij}$
  - $KTS6_{ij} = \gamma_{00} + u_{0j} + r_{ij}$
  - $KTS7_{ij} = \gamma_{00} + u_{0j} + r_{ij}$
  - $KTS8_{ij} = \gamma_{00} + u_{0j} + r_{ij}$
  - $KTS9_{ij} = \gamma_{00} + u_{0j} + r_{ij}$
  - $KTS10_{ij} = \gamma_{00} + u_{0j} + r_{ij}$
  - $UofL11_{ij} = \gamma_{00} + u_{0j} + r_{ij}$
-



Investigating teacher candidates nested in programs used two-level models using Hierarchical Linear Modeling 6.03 for analysis (Raudenbush, Bryk, Cheong, & Congdon, 2000). The results of the null models are provided in Table 18. The output tables are provided in Appendix C. To determine the proportion of variance that was attributed to between program variability and address the research question regarding reliability, the intraclass correlation coefficient (ICC) was calculated for each of the null models. The ICCs are provided in Table 19.

Table 18

*One-Way Random Effects ANOVA for Outcomes*

|             | Coefficient      |               |          | $\tau_{00}$ | Variability |          | <i>P</i> |
|-------------|------------------|---------------|----------|-------------|-------------|----------|----------|
|             | Coefficient (SE) | <i>t</i> (df) | <i>p</i> |             | $\sigma^2$  | $\chi^2$ |          |
| Inquiry     | 2.73 (.04)       | 74.62 (37)    | <.001    | .029        | .17         | 125.22   | <.001    |
| Action      | 2.67 (.04)       | 76.18 (37)    | <.001    | .024        | .20         | 101.58   | <.001    |
| Advocacy    | 2.64 (.04)       | 69.59 (37)    | <.001    | .029        | .20         | 112.80   | <.001    |
| Inquiry     | 2.73 (.04)       | 747.61 (37)   | <.001    | .026        | .18         | 110.40   | <.001    |
| Disposition |                  |               |          |             |             |          |          |
| Action      | 2.73 (.04)       | 82.812 (37)   | <.001    | .02         | .18         | 93.55    | <.001    |
| Disposition |                  |               |          |             |             |          |          |
| Advocacy    | 2.68 (.03)       | 77.59 (37)    | <.001    | .022        | .20         | 94.70    | <.001    |
| Disposition |                  |               |          |             |             |          |          |
| Impact      | 2.79 (.02)       | 116.39 (37)   | <.001    | .005        | .19         | 49.98    | .075     |
| Cycle 1     | 2.70 (.04)       | 60.89 (37)    | <.001    | .051        | .17         | 191.46   | <.001    |
| Cycle 2     | 2.78 (.04)       | 79.40 (37)    | <.001    | .028        | .15         | 130.03   | <.001    |
| Cycle 3     | 2.85 (.02)       | 129.75 (37)   | <.001    | .006        | .12         | 57.36    | .017     |
| Cycle 4     | 2.93 (.01)       | 205.78 (37)   | <.001    | .002        | .06         | 46.30    | .141     |
| KTS 1       | 2.80 (.03)       | 87.28 (37)    | <.001    | .022        | .13         | 117.04   | <.001    |
| KTS 2       | 2.74 (.03)       | 82.53 (37)    | <.001    | .021        | .17         | 97.11    | <.001    |
| KTS 3       | 2.79 (.03)       | 93.55 (37)    | <.001    | .018        | .14         | 97.58    | <.001    |
| KTS 4       | 2.78 (.03)       | 92.89 (37)    | <.001    | .017        | .14         | 100.03   | <.001    |
| KTS 5       | 2.63 (.04)       | 63.28 (37)    | <.001    | .040        | .20         | 134.68   | <.001    |
| KTS 6       | 2.65 (.04)       | 65.25 (37)    | <.001    | .039        | .18         | 138.09   | <.001    |
| KTS 7       | 2.73 (.03)       | 82.04 (37)    | <.001    | .021        | .18         | 97.07    | <.001    |
| KTS 8       | 2.74 (.03)       | 86.03 (37)    | <.001    | .019        | .17         | 90.90    | <.001    |
| KTS 9       | 2.78 (.03)       | 82.22 (37)    | <.001    | .025        | .15         | 119.54   | <.001    |
| KTS 10      | 2.72 (.03)       | 75.20 (37)    | <.001    | .028        | .18         | 116.88   | <.001    |
| UofL 11     | 2.69 (.04)       | 73.99 (37)    | <.001    | .028        | .18         | 119.24   | <.001    |

The Inquiry outcome assessed in the Ideas to Action Holistic Construct Rubric (Unit Key Assessment 7) at exit of the program aligns with the definition of Inquiry in the CEHD's conceptual framework aligned with critical thinking (2008). The first null model, shown in Table 18 shows that there was statistically significant program level variability in the assessment of Inquiry ( $p < .001$ ). The null model revealed that 14.47% of the variability in Inquiry could be attributed to between-program differences (intraclass correlation coefficient [ICC] = .145).

Table 18 provides the results of the HLM analysis for the Action outcome. Action is also captured in the Ideas to Action Holistic Construct Rubric (Unit Key Assessment 7) and the rubric language is reflective of the CEHD's conceptual framework aligned with problem solving (2008). The results of the null model show that there was statistically significant program level variability in the assessment of Action ( $p < .001$ ). The null model revealed that 10.69% of the variability in Inquiry could be attributed to between-program differences (intraclass correlation coefficient [ICC] = .107).

Advocacy is the third and final row of the Ideas to Action Holistic Construct Rubric (Unit Key Assessment 7) and the language within the row explicitly reflects the language of the CEHD's conceptual framework aligned with professional leadership and a commitment to social justice and equity. The null model for Advocacy, shown in Table 18 shows that there was statistically significant program level variability in the assessment of Inquiry ( $p < .001$ ). The null model revealed that 12.86% of the variability in Advocacy could be attributed to between-program differences (intraclass correlation coefficient [ICC] = .129).

The Ideas to Action Unit Dispositions Rubric (Unit Key Assessment 8) assesses Inquiry Disposition, Action Disposition, and Advocacy Disposition aligned with the CEHD's conceptual framework (2008). Inquiry Disposition assesses the teacher candidates disposition to inform practice through inquiry and reflection (CEHD's conceptual framework, 2008). Analysis of the null model (Table 18) revealed statistically significant program level variability in the assessment of Inquiry Disposition ( $p < .001$ ). Further, 12.8% of the variability in Inquiry Disposition could be attributed to between-program differences (intraclass correlation coefficient [ICC] = .128).

The outcome, Action Disposition, was designed to assess a teacher candidate's disposition to critique and change practice through content, pedagogical, and professional knowledge (CEHD's conceptual framework, 2008). The null model for Action Disposition revealed statistically significant variability ( $p < .001$ ) at the program level (Table 18). The intraclass correlation coefficient (ICC) was .101 and revealed that 10.08% of the variance in Action Disposition by programs could be attributed to between program differences.

Advocacy Disposition assesses a candidate's disposition "to affirm principles of social justice and equity and a commitment to making a positive difference" (CEHD Conceptual Framework, 2008). This is the third row of the Ideas to Action Unit Dispositions rubric (Unit Key Assessment 8). The null model for Advocacy Disposition as the outcome variable revealed that there was statistically significant variability ( $p < .001$ ) in Advocacy Disposition at the program level and that 10.24% of that variance could be attributed to between program differences (intraclass correlation coefficient [ICC] = .102).

Impact assessment measures the teacher candidate's analysis of student learning and evidence of impact on P-12 student learning. The Impact assessment is included in the Instructional Sequence assessment that is completed by teacher candidates within the student teaching semester. The null model for the Impact assessment revealed that there was not statistically significant program variability ( $p=.075$ ) at the .05 alpha level for the Impact assessment.

Cycle 1 through Cycle 4 capture the holistic assessments of the teacher candidate's performance during each of the four formal observations completed by the university supervisor during the student teaching semester. The rubric used to capture this assessment addresses Kentucky Teacher Standards (KTS) 1, 2, 3, 4, 5, 6, 7, and University of Louisville's Diversity Standard 11. The results revealed that there was statistically significant program variability for Cycle 1 ( $p<.001$ ) and Cycle 2 ( $p<.001$ ). The ICC for Cycle 1 was .235 and the ICC for Cycle 2 was .158, revealing that 23.5% of the program variability in Cycle 1 outcome could be attributed to between program variability and 15.78% of the program variability in Cycle 2 outcome could be attributed to between program variability. Cycle 3 results revealed statistically significant program variability ( $p=.017$ ), with an ICC of .049. This indicates that 4.92% of the variability by program could be attributed to between program variability. The null model with Cycle 4 as outcome revealed that there was no statistically significant program variability ( $p=.141$ ) in Cycle 4.

Table 18 provides the results for the null models with each of the Kentucky Teacher Standards (KTS) and the University of Louisville Diversity Standard 11 as the outcomes. All of these standards are assessed by faculty at the completion of the

candidate's exit work sample or exit portfolio. For each standard, there was statistically significant program variability ( $p < .001$ ) in each of the standards.

The ICC for all outcomes tested are included in Table 19. Cycle 1 had the highest proportion of between program variability (23.47%). All of the ICCs reveal small proportion of between program variability as compared to within program variability; however, further contextual analysis was conducted to investigate the between program variability.

Table 19

*Intraclass Correlation Coefficients (ICCs)*

| <b>Milestone</b>     | <b>ICC</b> | <b>Proportion of Variance Between Program</b> |
|----------------------|------------|---|
| Inquiry              | .145       | 14.47%  |
| Action               | .107       | 10.69%  |
| Advocacy             | .129       | 12.86%  |
| Inquiry Disposition  | .128       | 12.80%  |
| Action Disposition   | .101       | 10.08%  |
| Advocacy Disposition | .102       | 10.24%  |
| Impact               | .025       | 2.46%   |
| Cycle1               | .235       | 23.47%  |
| Cycle2               | .158       | 15.78%  |
| Cycle3               | .049       | 4.92%   |
| Cycle4               | .027       | 2.68%   |
| KTS1                 | .144       | 14.36%  |
| KTS2                 | .111       | 11.14%  |
| KTS3                 | .113       | 11.33%  |
| KTS4                 | .105       | 10.51%  |
| KTS5                 | .170       | 17.03%  |
| KTS6                 | .173       | 17.30%  |
| KTS7                 | .108       | 10.83%  |
| KTS8                 | .098       | 9.83%   |
| KTS9                 | .139       | 13.89%  |
| KTS10                | .133       | 13.30%  |
| UofL11               | .136       | 13.56%  |

To further investigate the between-program variability, level 2 variables were added to the null model to run a contextual model to determine if degree type (bachelors or masters) or program type (elementary, middle and secondary, or special education) had a statistically significant impact on the outcomes. The equations for this model are shown in Table 20.

Table 20

*Contextual Models*

| <b>Contextual Model for Program Type</b>  | <b>Contextual Model for Degree Type</b>   |
|---|---|
| $INQUIRY_{ij} = \gamma_{00} + \gamma_{01} * PROGTYPE_j + u_{0j} + r_{ij}$             | $INQUIRY_{ij} = \gamma_{00} + \gamma_{01} * DEGREETYPE_j + u_{0j} + r_{ij}$             |
| $ACTION_{ij} = \gamma_{00} + \gamma_{01} * PROGTYPE_j + u_{0j} + r_{ij}$              | $ACTION_{ij} = \gamma_{00} + \gamma_{01} * DEGREETYPE_j + u_{0j} + r_{ij}$              |
| $ADVOCACY_{ij} = \gamma_{00} + \gamma_{01} * PROGTYPE_j + u_{0j} + r_{ij}$            | $ADVOCACY_{ij} = \gamma_{00} + \gamma_{01} * DEGREETYPE_j + u_{0j} + r_{ij}$            |
| $INQUIRYDISPOSITION_{ij} = \gamma_{00} + \gamma_{01} * PROGTYPE_j + u_{0j} + r_{ij}$  | $INQUIRYDISPOSITION_{ij} = \gamma_{00} + \gamma_{01} * DEGREETYPE_j + u_{0j} + r_{ij}$  |
| $ACTIONDISPOSITION_{ij} = \gamma_{00} + \gamma_{01} * PROGTYPE_j + u_{0j} + r_{ij}$   | $ACTIONDISPOSITION_{ij} = \gamma_{00} + \gamma_{01} * DEGREETYPE_j + u_{0j} + r_{ij}$   |
| $ADVOCACYDISPOSITION_{ij} = \gamma_{00} + \gamma_{01} * PROGTYPE_j + u_{0j} + r_{ij}$ | $ADVOCACYDISPOSITION_{ij} = \gamma_{00} + \gamma_{01} * DEGREETYPE_j + u_{0j} + r_{ij}$ |
| $CYLCE1_{ij} = \gamma_{00} + \gamma_{01} * PROGTYPE_j + u_{0j} + r_{ij}$              | $CYLCE1_{ij} = \gamma_{00} + \gamma_{01} * DEGREETYPE_j + u_{0j} + r_{ij}$              |
| $CYLCE2_{ij} = \gamma_{00} + \gamma_{01} * PROGTYPE_j + u_{0j} + r_{ij}$              | $CYLCE2_{ij} = \gamma_{00} + \gamma_{01} * DEGREETYPE_j + u_{0j} + r_{ij}$              |
| $CYLCE3_{ij} = \gamma_{00} + \gamma_{01} * PROGTYPE_j + u_{0j} + r_{ij}$              | $CYLCE3_{ij} = \gamma_{00} + \gamma_{01} * DEGREETYPE_j + u_{0j} + r_{ij}$              |
| $KTS1_{ij} = \gamma_{00} + \gamma_{01} * PROGTYPE_j + u_{0j} + r_{ij}$                | $KTS1_{ij} = \gamma_{00} + \gamma_{01} * DEGREETYPE_j + u_{0j} + r_{ij}$                |
| $KTS2_{ij} = \gamma_{00} + \gamma_{01} * PROGTYPE_j + u_{0j} + r_{ij}$                | $KTS2_{ij} = \gamma_{00} + \gamma_{01} * DEGREETYPE_j + u_{0j} + r_{ij}$                |
| $KTS3_{ij} = \gamma_{00} + \gamma_{01} * PROGTYPE_j + u_{0j} + r_{ij}$                | $KTS3_{ij} = \gamma_{00} + \gamma_{01} * DEGREETYPE_j + u_{0j} + r_{ij}$                |
| $KTS4_{ij} = \gamma_{00} + \gamma_{01} * PROGTYPE_j + u_{0j} + r_{ij}$                | $KTS4_{ij} = \gamma_{00} + \gamma_{01} * DEGREETYPE_j + u_{0j} + r_{ij}$                |
| $KTS5_{ij} = \gamma_{00} + \gamma_{01} * PROGTYPE_j + u_{0j} + r_{ij}$                | $KTS5_{ij} = \gamma_{00} + \gamma_{01} * DEGREETYPE_j + u_{0j} + r_{ij}$                |
| $KTS6_{ij} = \gamma_{00} + \gamma_{01} * PROGTYPE_j + u_{0j} + r_{ij}$                | $KTS6_{ij} = \gamma_{00} + \gamma_{01} * DEGREETYPE_j + u_{0j} + r_{ij}$                |
| $KTS7_{ij} = \gamma_{00} + \gamma_{01} * PROGTYPE_j + u_{0j} + r_{ij}$                | $KTS7_{ij} = \gamma_{00} + \gamma_{01} * DEGREETYPE_j + u_{0j} + r_{ij}$                |
| $KTS8_{ij} = \gamma_{00} + \gamma_{01} * PROGTYPE_j + u_{0j} + r_{ij}$                | $KTS8_{ij} = \gamma_{00} + \gamma_{01} * DEGREETYPE_j + u_{0j} + r_{ij}$                |
| $KTS9_{ij} = \gamma_{00} + \gamma_{01} * PROGTYPE_j + u_{0j} + r_{ij}$                | $KTS9_{ij} = \gamma_{00} + \gamma_{01} * DEGREETYPE_j + u_{0j} + r_{ij}$                |
| $KTS10_{ij} = \gamma_{00} + \gamma_{01} * PROGTYPE_j + u_{0j} + r_{ij}$               | $KTS10_{ij} = \gamma_{00} + \gamma_{01} * DEGREETYPE_j + u_{0j} + r_{ij}$               |
| $UofL11_{ij} = \gamma_{00} + \gamma_{01} * PROGTYPE_j + u_{0j} + r_{ij}$              | $UofL11_{ij} = \gamma_{00} + \gamma_{01} * DEGREETYPE_j + u_{0j} + r_{ij}$              |

The analysis revealed that there were several instances in which degree type or program type had a statistically significant ( $\alpha = .05$ ) impact on the outcome. The between program variance for the null model was compared to the between program variance in the contextual model. The percentage of variance in between program differences on the outcomes, accounted for by degree type or program type are included in Table 22. The effect of degree type on Inquiry ( $p=.029$ ), Advocacy ( $p=.007$ ), Inquiry Disposition ( $p=.030$ ), Advocacy Disposition ( $p=.040$ ), Cycle 1 ( $p<.001$ ), Cycle 2 ( $p<.001$ ), KTS 1 ( $p=.047$ ), and KTS4 ( $p=.024$ ), was statistically significant at the .05 alpha level. The effect of program type on KTS 2 ( $p=.018$ ), KTS 6 ( $p=.033$ ), KTS 8 ( $p=.032$ ), and KTS 10 ( $p<.001$ ) was statistically significant at the .05 alpha level.

Table 21

*Percentage of Between Program Variance Explained in Contextual Models*

| <b>Milestone</b>    | <b>DegreeType</b> | <b>ProgramType</b> |
|---------------------|-------------------|--------------------|
| C3INQUIRY           | <b>*15.30%</b>    | 0.17%              |
| C3ACTION            | 12.74%            | 10.75%             |
| C3ADVOCACY          | <b>*36.34%</b>    | -3.29%             |
| InquiryDisposition  | <b>*19.25%</b>    | 7.08%              |
| ActionDisposition   | 14.41%            | 12.78%             |
| AdvocacyDisposition | <b>*19.67%</b>    | -1.43%             |
| Impact              | --                | --                 |
| Cycle1              | <b>*33.74%</b>    | 3.15%              |
| Cycle2              | <b>*36.83%</b>    | -2.45%             |
| Cycle3              | 6.60%             | 0.16%              |
| Cycle4              | --                | --                 |
| KTS1                | <b>*15.58%</b>    | 14.24%             |
| KTS2                | 0.98%             | <b>*22.24%</b>     |
| KTS3                | 2.97%             | 18.11%             |
| KTS4                | <b>*21.66%</b>    | 19.78%             |
| KTS5                | 2.47%             | 2.42%              |
| KTS6                | -4.33%            | <b>*18.44%</b>     |
| KTS7                | 8.28%             | 10.29%             |
| KTS8                | 1.23%             | <b>*25.27%</b>     |
| KTS9                | -2.40%            | 13.36%             |

Table 21 Continued

|        |        |                |
|--------|--------|----------------|
| KTS10  | -4.64% | <b>*48.44%</b> |
| UofL11 | 5.50%  | 15.26%         |

\* indicates  $p < .05$

The results of the contextual models reveal that 15.30% of the between program variance in Inquiry, 36.64% of the between program in Advocacy, 19.25% of the between program variance in Inquiry Disposition, 19.67% of the between program variance in Advocacy Disposition, 33.74% of the between program variance in Cycle 1, 36.83% of the between program variance in Cycle 2, 15.58% of the between program variability in KTS 1, and 21.66% of the between program variance in KTS 4 can be attributed to whether or not the program is a B.S. or MAT program. Further, 22.24% of the between program variance in KTS 2, 18.44% of the between program variance in KTS 6, 25.27% of the between program variance in KTS 8, and 48.44% of the between program variance in KTS 10 can be attributed to whether the program is elementary, middle and secondary, or special education. The contextual analysis reveals that degree type (B.S. or MAT) or program type (Elementary, Middle/Secondary, or Special Education) can have significant impacts on variability of ratings across programs. This may be due to varying teacher candidate knowledge, skills, and dispositions prior to admittance to programs or variations in preparation program requirements. The findings of the contextual analysis was incorporated into the discussion and next steps provided in Chapter V.

### **Validity Analysis**

The fourth and final research question focuses on the validity of assessments used to measure knowledge, skills, and dispositions in teacher education programs. For this



metaevaluation, the focus was specifically on the construct validity of measures within the CEHD’s assessment system aligned with the CEHD’s conceptual framework. The conceptual framework for the University of Louisville’s CEHD, *Shaping Tomorrow: Ideas to Action*, focuses on preparing CEHD students to be critical thinkers, problem solvers, and professional leaders. The conceptual framework includes three constructs of Inquiry, Action, and Advocacy. Table 22 provides the constructs, the constructs as learned and applied, the constructs as qualities reflected in teacher candidates and graduates, and then the dispositions that are reflected in teacher candidates and graduates.

Table 22

*CEHD Conceptual Framework Aligned with Candidate Knowledge, Skills, and Dispositions (2007, p. 17)*

| Conceptual Framework Constructs           | Inquiry  | Action   | Advocacy   |
|---|--|--|--|
| Constructs as Learned and Applied         | Research   | Practice   | Service  |
| Constructs Reflected in Candidates        | Critical Thinkers  | Problem Solvers  | Professional Leaders   |
| Unit Dispositions Reflected in Candidates | Exhibits a disposition to inform practice through inquiry and reflection | Exhibits a disposition to critique and change practice through content, pedagogical, and professional knowledge. | Exhibits a disposition to affirm principles of social justice and equity and a commitment to making a positive difference. |

The CEHD embedded the conceptual framework into the continuous assessment system for tracking teacher candidate performance data throughout teacher education

programs by strategically mapping the conceptual framework constructs to the items measured within the 10 Unit Key Assessment rubrics. Each item is embedded within a row of a rubric, and teacher candidates are scored on the scale of Target (3), Acceptable (2), and Unacceptable (1). These assessments were designed to reflect the conceptual framework and often include language taken directly from the conceptual framework document. The items included in the analysis for construct validity are provided in Table 23. This study focuses on these assessments and how well they measure the constructs as reflected in candidates, which are critical thinking, problem solving, and professional leadership. The rubrics used to assess each of the outcomes were developed by faculty of CEHD teacher education programs, in partnership with P-12 school partners, and field and clinical staff. Teacher candidates submit assessment artifacts through LiveText®, and faculty complete the assigned assessment rubric electronically to provide feedback to the teacher candidate on his or her performance and capture data on outcomes.

Table 23

*Model Constructs, Measured Items, and Item Description*

| Construct                | Items                 | Item Description   |
|--------------------------|-----------------------|--|
| <b>Critical Thinking</b> | Inquiry               | Rubric Row: Inquiry "... through active engagement and skilled training in multiple methods of rigorous Research candidates in the CEHD develop the knowledge, skills, and dispositions to become Critical Thinkers/" (CF, p.18) |
|                          | Inquiry Disposition   | Rubric Row: Exhibits a disposition to inform practice through inquiry and reflection   |
|                          | Content Knowledge     | Rubric Row: Kentucky Teaching Standard 1: Demonstrates Applied Content Knowledge   |
|                          | Reflection Evaluation | Rubric Row: Kentucky Teaching Standard 7: Reflects On and Evaluates Teaching and Learning<br>Rubric Row: Kentucky Teaching Standard 9: Evaluates Teaching and Implements Professional Development                                |

Table 23 Continued

|                                |                                   |  |
|--------------------------------|-----------------------------------|--|
| <b>Problem Solving Ability</b> | Action                            | Rubric Row: Action "... through routine, continual, and pervasive Practice—whether this be in the areas of pedagogy and instructional leadership, counseling, or research—candidates in the CEHD develop the knowledge, skills, and dispositions to become Problem Solvers in the community” (CF, p. 19) |
|                                | Action Disposition                | Rubric Row: Exhibits a disposition to critique and change practice through content, pedagogical, and professional knowledge  |
|                                | Planning                          | Rubric Row: Kentucky Teaching Standard 2: Designs and Plans Instruction  |
|                                | Classroom Management              | Rubric Row: Kentucky Teaching Standard 3: Creates and Maintains Learning Climate   |
|                                | Implements Instruction Technology | Rubric Row: Kentucky Teaching Standard 4: Implements and Manages Instruction<br>Rubric Row: Kentucky Teaching Standard 6: Demonstrates the Implementation of Technology  |
| <b>Professional Leadership</b> | Advocacy                          | Rubric Row: Advocacy "... through dedicated, committed Service to their community, and world candidates in the CEHD develop the knowledge, skills, and dispositions to become Professional Leaders” (CF, p. 20)  |
|                                | Advocacy Disposition              | Rubric Row: Exhibits a disposition to affirm principles of social justice and equity and a commitment to making a positive difference  |
|                                | Impact Diversity                  | Rubric Row: Impact on P-12 Student Learning<br>Rubric Row: University of Louisville Standard 11: Understands the Complex Lives of Students and Adults in Schools and Society   |
|                                | Assessment                        | Rubric Row: Kentucky Teaching Standard 5: Assesses and Communicates Learning Results   |
|                                | Collaboration                     | Rubric Row: Kentucky Teaching Standard 8: Collaborates with Colleagues/Parents/Others  |
|                                | Leadership Standard               | Rubric Row: Kentucky Teaching Standard 10: Provides leadership within school/community/profession  |

For the purpose of this study, critical thinking, problem solving, and professional leadership are the model constructs in a confirmatory factor analysis to determine how well the measures set in place are measuring the constructs. Table 23 provides an outline of the items that are being used to measure each of the model constructs. Figure 2,

provides the base model to be analyzed using Structural Equation modeling in SPSS AMOS software. The sample for CFA includes all teacher education program completers from the 2011-2012 and 2012-2013 academic years (Summer 2011-Spring 2013). Each of the items included in the model are measured on a 3-point scale (3=Target, 2=Acceptable, 1=Unacceptable).

As shown in Figure 2, there are five items used to measure critical thinking ability in teacher candidates in CEHD teacher education programs. The items include an assessment of Inquiry, Disposition related to Inquiry, demonstrated Content Knowledge (Kentucky Teacher Standard 1), ability to Reflect on and Evaluate Teaching (Kentucky Teacher Standard 7), and ability to Evaluate Teaching and Implement Professional Development (Kentucky Teacher Standard 9). The Inquiry assessment is a direct reflection of the CEHD's conceptual framework as Inquiry is the construct associated with critical thinking. The Inquiry Disposition assessment is also directly aligned with the conceptual framework, as the rubric assessment is defined by the conceptual framework description of the disposition for that construct. The Content Knowledge (KTS 1) assessment is a direct reflection of a teacher candidate's ability to attain information specific to his or her professional program and ability to communicate their own understanding through assessment artifacts and work samples. As defined by the CEHD's conceptual framework, critical thinking is associated with reflection and questioning, which supports the faculty's decision to associate reflection on teaching (KTS 7 and KTS 9) with critical thinking.

Problem solving is measured with how well a teacher candidate demonstrates application of knowledge with the Action assessment and how well a teacher candidate

exhibits a disposition that focuses on improving education through the application of their knowledge with the Action Disposition assessment. Further, Kentucky Teacher Standards (KTS) 2, 3, 4, and 6 are focused on the application of knowledge and skills in the classroom setting, and, therefore, support the faculty's alignment of these items with the problem solving.

Professional Leadership is measured through seven assessment points. They include Advocacy, Advocacy Disposition, Impact, Diversity (UofL Standard 11), Assessment (KTS 5), Collaboration (KTS 8), and Leadership Standard (KTS 10). Advocacy is directly assessed through assessment artifacts that demonstrate a commitment to serving the community. The Advocacy Disposition assessment focuses on a teacher candidate's ability to portray a personal disposition towards equity and making a difference as an educator. The Impact assessment is a reflection on teaching and understanding of how the teacher candidate is making a positive difference in the lives of P-12 students. Kentucky Teacher Standard 5 ensures that teacher candidates assess and communicate the results of P-12 student learning. Kentucky Teacher Standard 8 focuses on collaborating with colleagues, parents, and other to support P-12 student learning, and Kentucky Teaching Standard 10 assessment specifically evaluates teacher candidate assessment artifacts related to addressing leadership in the school and community setting. Ultimately, all of these assessments are tied back to being a professional leader in the school and community to promote learning of students and the greater community.

***Base Model (Model 1).*** For the base model (Figure 3), critical thinking, problem solving, and professional leadership serve as latent constructs (exogenous variables) and are all correlated with one another because of the connections defined by the CEHD's

conceptual framework (2008). Additionally, Inquiry, Action, and Advocacy errors, as well as Inquiry Disposition, Action Disposition, and Advocacy Disposition are all correlated because they are assessed within the same rubric. Inquiry, Action, and Advocacy were identified as marker variables as they are most closely aligned with the latent constructs as reflected in teacher candidates.

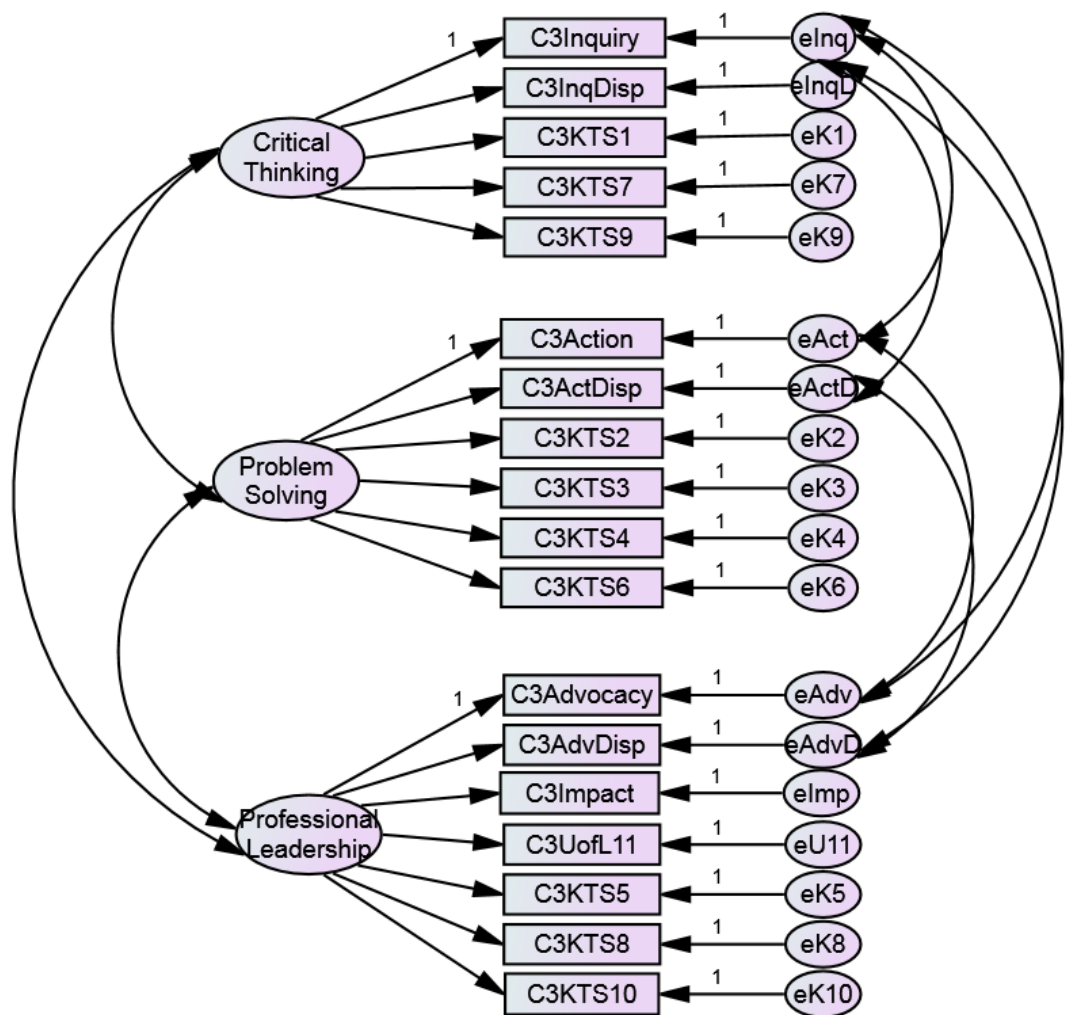


Figure 3. Confirmatory Factor Analysis Base Model. Critical Thinking, Problem Solving, and Professional Leadership are all correlated as exogenous variables, which are all

interrelated due to the relationships defined in the conceptual framework. The conceptual framework constructs of Inquiry, Action, and Advocacy are all correlated because all three items are embedded within the same rubric. The errors of Inquiry Disposition, Action Disposition, and Advocacy Disposition are also correlated since they are assessed in the same rubric. The endogenous variables associated with Critical Thinking include C3Inquiry (Inquiry), C3InqDisp (Inquiry Disposition), C3KTS1 (KTS 1 – Demonstrates Applied Content Knowledge), C3KTS7 (KTS 7 – Reflects on and Evaluates Teaching and Learning), and C3KTS9 (KTS 9 – Evaluates Teaching and Implements Professional Development). The endogenous variables associated with Problem Solving include C3Action (Action), C3ActDisp (Action Disposition), C3KTS2 (KTS 2 – Designs and Plans Instruction), C3KTS3 (KTS 3 – Creates and Maintains Learning Climate), C3KTS4 (KTS 4 – Implements and Manages Instruction), and C3KTS6 (KTS 6 – Demonstrates the Implementation of Technology). The endogenous variables associated with Professional Leadership include C3Advocacy (Advocacy), C3AdvDisp (Advocacy Disposition), C3Impact (Impact on P-12 Student Learning), C3UofL11 (UofL Diversity Standard), C3KTS5 (KTS 5 – Assesses and Communicates Learning Results), C3KTS8 (KTS 8 – Collaborates with Colleagues/Parents/Others), and C3KTS10 (KTS 10 – Provides Leadership within School/Community/Profession).

The initial analysis of the base model resulted in poor model fit using multiple model fit statistics. The TLI was .807 which was below the acceptable range as suggested by Bentler & Bonett (1980), and the CLI was .841 which was also below the acceptable range as suggested by Hu & Bentler (1999). The  $\chi^2$  was 979.93 (126),  $p < .001$  and therefore statistically significant suggesting poor model fit (Kline, 2011). The RMSEA

was .119 which above the recommended upper limit of .08 for acceptable model fit as suggested by McCallum, Browne, and Sugawara (1996).

In the base model, all paths and correlations were statistically significant and, therefore, no paths or correlations were removed from the model. Further, there were no Heywood cases (standardized loadings greater than 1 or negative error variances) and, therefore, the results were interpretable. A review of the standardized residual covariances revealed several values greater than the recommended value of  $\pm 1.96$ . Further the modification indices in the base model suggested that there was a relationship between the errors of KTS8 and KTS10 (M.I. = 97.502), the errors of KTS 8 and KTS 9 (M.I. = 93.745), the errors of KTS 9 and KTS 10 (M.I. = 89.435, and the errors of Advocacy and Advocacy Disposition (M.I. 55.780).

**Model 2.** Due to the large modification index of 97.502 for the errors of KTS 8 and KTS 10, as well as a standardized residual covariance between those two items of 7.305, the second model was run with a correlation added between KTS 8 and 10. This modification is supported by theory as standards 8, 9, and 10 are primarily assessed based on the teacher candidate's professional development plans developed during the student teaching semester. As shown in Table 24, the model fit statistics for model 2 (Figure 4) did improve slightly from the base model; however, they still suggested poor model fit. A  $\chi^2$  difference test between the base model and model 2 revealed that  $\chi^2 = 107.853(1)$ ,  $p < .001$ . Therefore, model 2 was preferred.



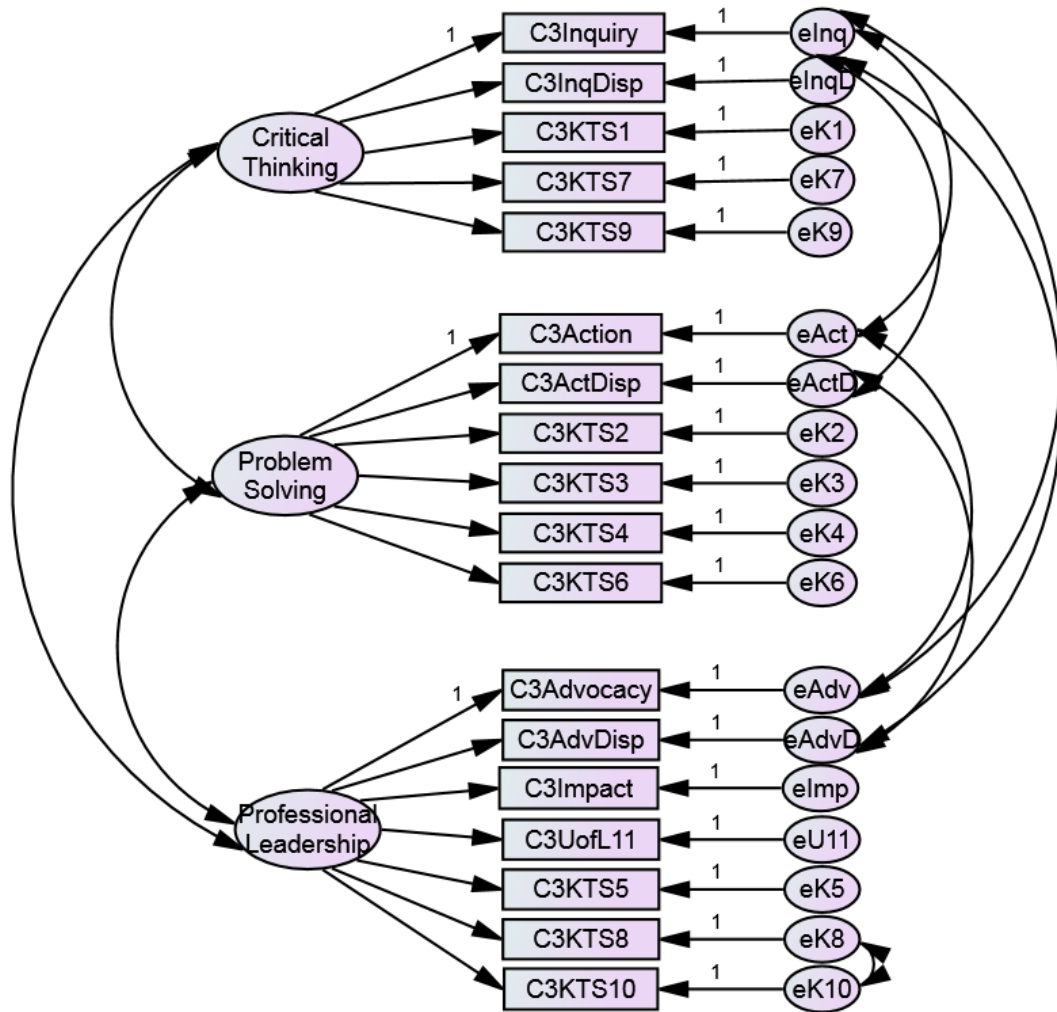


Figure 4. Model 2 for Confirmatory Factor Analysis (CFA). Correlation added between the errors of Kentucky Teacher Standard 8 and Kentucky Teacher Standard 10.

**Model 3.** A review of standardized residual covariances from model 2 still revealed several values greater than  $\pm 1.96$ . KTS 9, and KTS 10 had the highest standardized residual covariance of 8.434. Further, the modification indices (M.I. = 36.133) suggested relationships between the errors of KTS 9 and KTS 10. The same theory that supported the addition of the correlation between KTS 8 and 10 also applied

to KTS 9 and 10. Therefore, model 3 (Figure 5) was run with a correlation added between the errors of KTS 9 and KTS 10. As shown in Table 24, the model fit statistics did improve slightly; however, they still suggest poor model fit. A  $\chi^2$  difference test between the model 2 and model 3 revealed that  $\chi^2 = 43.116(1)$ ,  $p < .001$ ; therefore, model 3 was preferred.

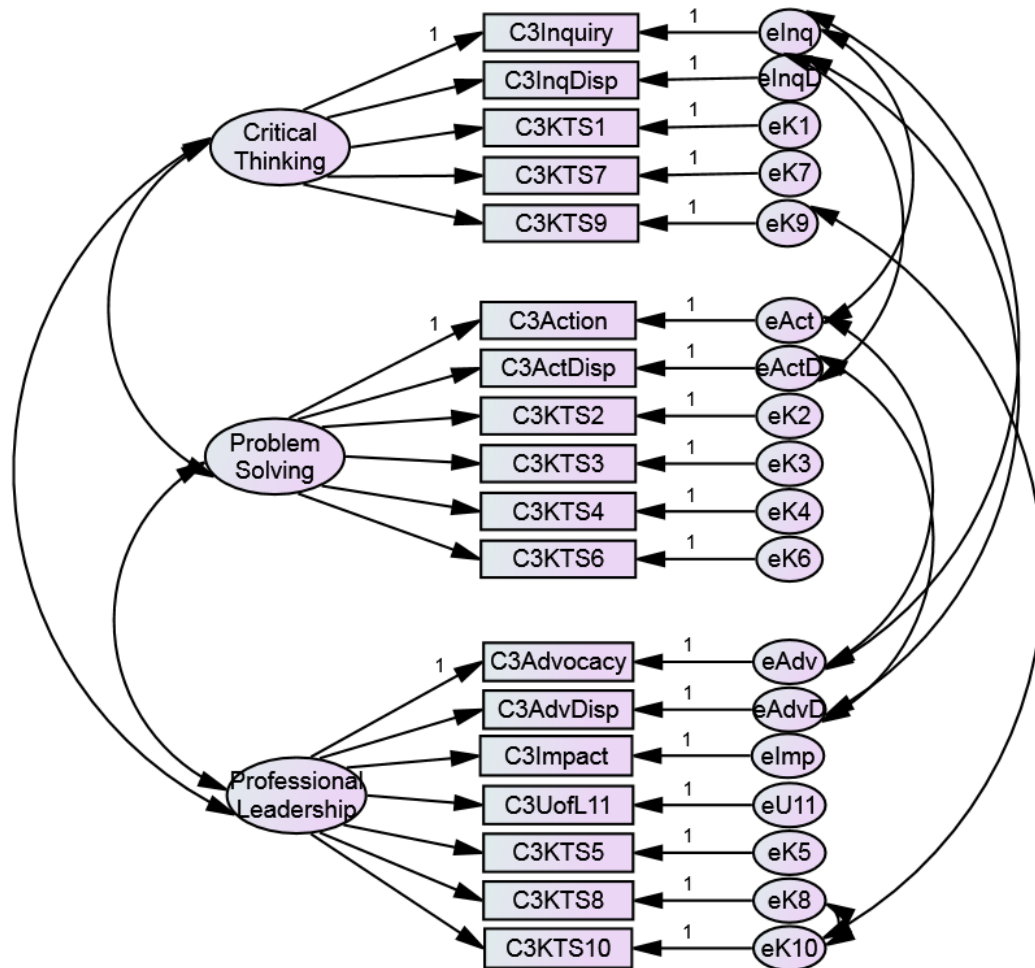


Figure 5. Model 3 for Confirmatory Factor Analysis (CFA). Correlation added between the errors of Kentucky Teacher Standard 9 and Kentucky Teacher Standard 10.

**Model 4.** Model 3 standardized residual covariance for KTS 8 and KTS 9 was 8.321, and the modification index for KTS 8 and 9 suggested a relationship between the errors of these two items (M.I. = 76.714). Supported by the theory that those standards assess the professional development plans that are developed by teacher candidates during the student teaching semester, a correlation was added between the errors of KTS 8 and KTS 9 for model 4 (Figure 6). A  $\chi^2$  difference test between the model 3 and model 4 revealed that  $\chi^2 = 116.517(1), p < .001$ ; therefore, model 4 was preferred over model 3.

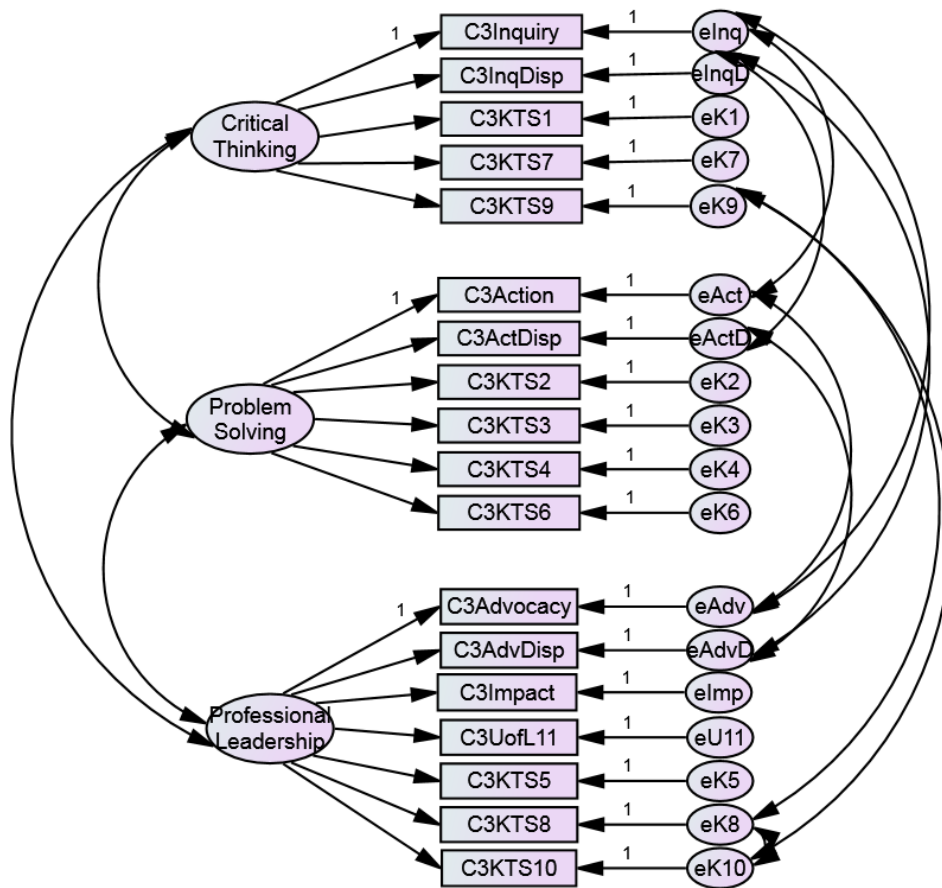


Figure 6. Model 4 for Confirmatory Factor Analysis (CFA). Correlation added between the error of Kentucky Teacher Standard 8 and Kentucky Teacher Standard 9.

**Model 5.** In Model 4, the standardized residual covariance between Advocacy and Advocacy disposition was greater than  $\pm 1.96$ , and the modification indices suggested a relationship between the errors of these two items (M.I. = 47.899). Theory supports the addition of a correlation between the errors of these two items for the reasoning that the definitions of these two items are closely aligned, as Advocacy is the CEHD's conceptual framework construct and Advocacy Disposition is the construct reflected in teacher candidates. As shown in Table 24, the model fit statistics for model 5 (Figure 7) have improved slightly with the additional correlation between the errors of Advocacy and Advocacy Disposition, suggesting acceptable model fit. The CFI (.903) was within the acceptable range as suggested by Hu & Bentler (1999); the TLI is nearing the acceptable range of .90 as suggested by Bentler and Bonett (1980); and the RMSEA was then within the marginal range as suggested by McCallum, Browne, and Sugawara (1996). The  $\chi^2$  difference test between the model 4 and model 5 revealed that  $\chi^2 = 69.908(1), p < .001$ ; therefore, model 5 was preferred over model 4.

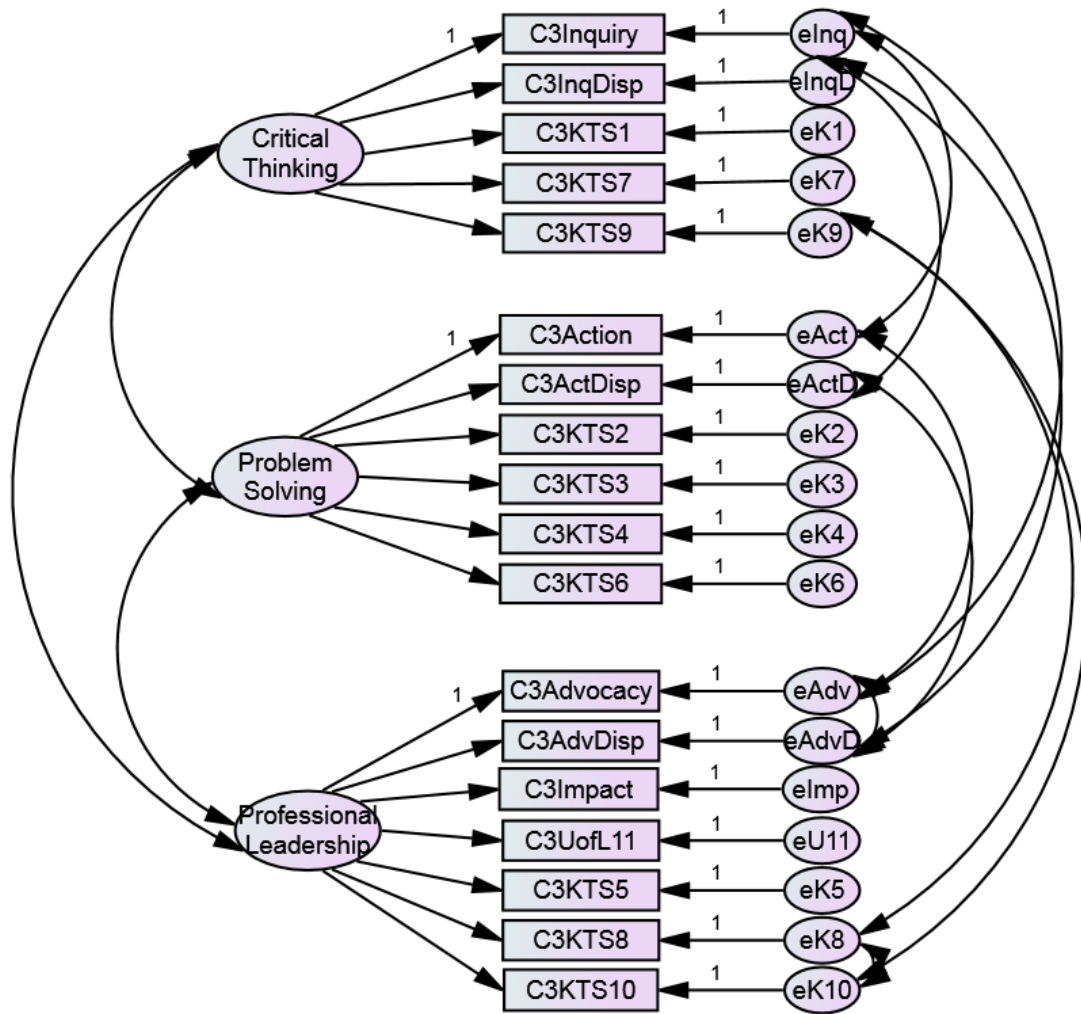


Figure 7. Model 5 for Confirmatory Factor Analysis (CFA). Correlation added between the errors of Advocacy and Advocacy Disposition.

In Model 5, the modification indices suggested an additional correlation between the errors of KTS 3 and KTS 4 (M.I. = 29.286). A correlation between the errors of these two items is supported by theory as a result of the connections between Creating and Maintaining a Learning Climate (KTS 3) and Implementing and Managing Instruction (KTS 4). A revised model was run with the added correlation between the errors of these

two items. The notes for the model stated that the model was not admissible, and, therefore, the results could not be interpreted. The error stated that the covariance matrix was not positive definite. This suggests that there may be an issue with multicollinearity in the model or discriminant validity. A review of the correlations from Model 5 revealed that the correlations between the three latent constructs of Critical Thinking, Problem Solving, and Professional Leadership are all highly correlated. Critical Thinking and Problem Solving were correlated at  $r=.971$ , Critical Thinking and Professional Leadership were correlated at  $r=.910$ , and Problem Solving and Professional Leadership were correlated at  $r=.975$ . As recommended by Kline (2011), correlations between factors should not be above .85, and the consequences may be multicollinearity and problems with convergence. Additional modification indices were applied to test the model with additional correlations for investigative purpose; however, all models resulted in the same not positive definite error.

**Model 6.** In model 5, there were no additional correlations that could be supported by theory and result in an interpretable model. Next, the standardized regression weights or loadings were reviewed. All loadings were above the recommended level of .40 (Stevens, 2009), except for the loading for Impact. For Model 6 (Figure 8), Impact was removed from the model. Model fit statistics shown in Table 24, represent marginal to acceptable model fit. Since model 5 and 6 are non-nested models, a  $\chi^2$  difference test could not be used to compare models. The AIC and BIC were used to compare the non-nested models. Table 25 shows the AIC and BIC are lower for model 6 than model 5. As suggested by Kline (2011), the lower value was the better model as it is more likely to replicate. Therefore, model 6 as shown in Figure 8, represents the final model for

measuring critical thinking, problem solving, and professional leadership. The  $\chi^2$  was still significant suggesting marginal model fit (Kline, 2011). The TLI value was near the acceptable range, and the CFI value suggests acceptable model fit (Bentler & Bonett, 1980; Hu & Bentler, 1999). The RMSEA was marginal (McCallum et al., 1996). Overall, the model was considered to have acceptable model fit.

Table 24

*Model Fit Statistics (RMSEA, TLI, CFI, and RMSEA)*

| <b>Model</b> | <b><math>\chi^2</math></b> | <b>TLI</b> | <b>CFI</b> | <b>RMSEA (90% CI)</b> |
|--------------|----------------------------|------------|------------|-----------------------|
| Base Model   | 979.934(126), $p < .001$   | .807       | .841       | .119 (.112-.126)      |
| Model 2      | 872.081(125), $p < .001$   | .830       | .861       | .111 (.104-.118)      |
| Model 3      | 828.965(124), $p < .001$   | .838       | .869       | .109 (.102-.116)      |
| Model 4      | 712.448(123), $p < .001$   | .864       | .890       | .100 (.093-.107)      |
| Model 5      | 642.540(122), $p < .001$   | .879       | .903       | .094 (.087-.101)      |
| Model 6      | 613.272(106), $p < .001$   | .877       | .904       | .100 (.092-.107)      |

Table 25

*Model 5 & 6: AIC and BIC Comparison*

| <b>Model</b> | <b>AIC</b> | <b>BIC</b> |
|--------------|------------|------------|
| Model 5      | 740.54     | 945.361    |
| Model 6      | 707.272    | 903.733    |

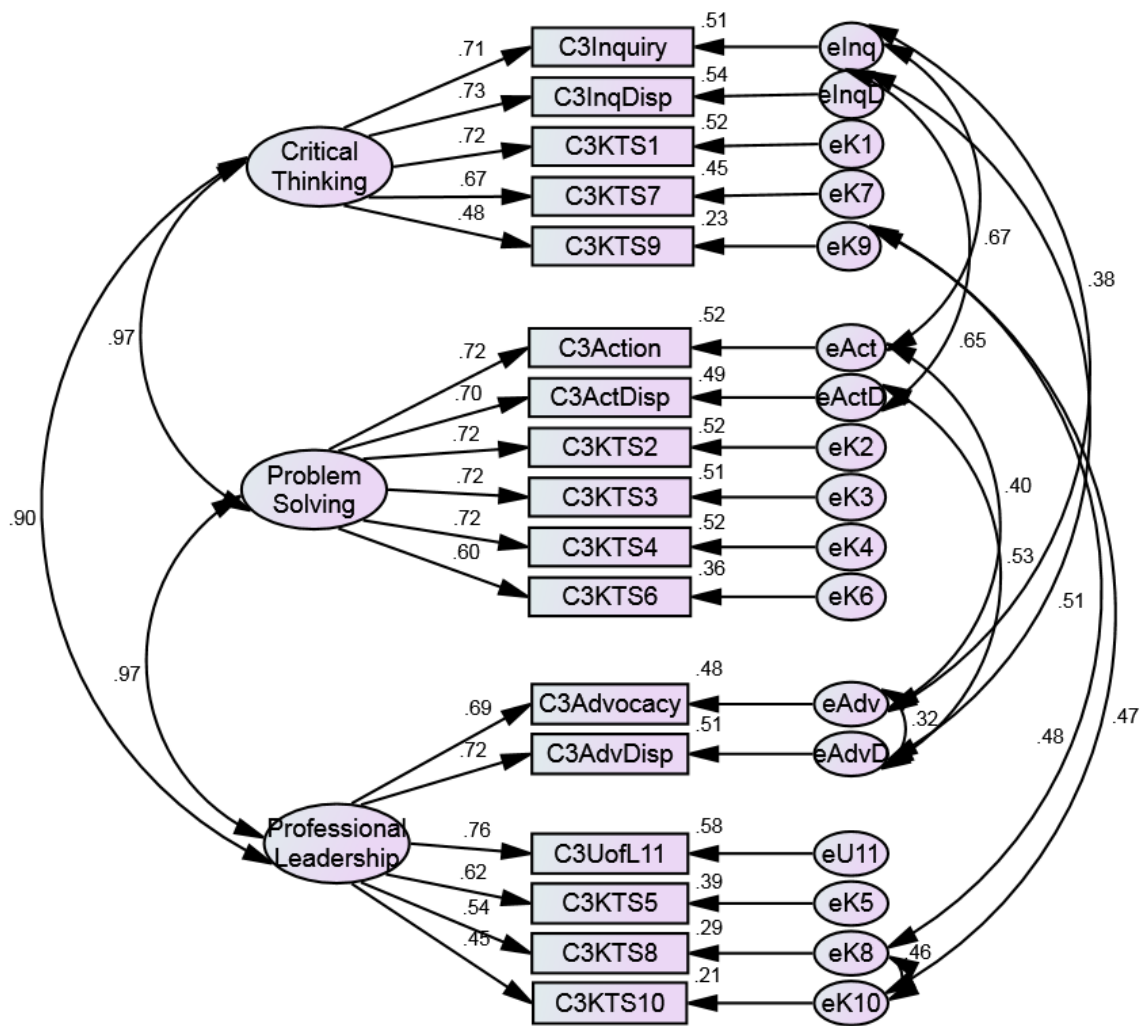


Figure 8. Final Model for Confirmatory Factor Analysis (CFA) with standardized coefficients. Impact removed from the model due to standardized loading below .40 (Stevens, 2009). The endogenous variables associated with Critical Thinking include C3Inquiry (Inquiry), C3InqDisp (Inquiry Disposition), C3KTS1 (KTS 1 – Demonstrates Applied Content Knowledge), C3KTS7 (KTS 7 – Reflects on and Evaluates Teaching and Learning), and C3KTS9 (KTS 9 – Evaluates Teaching and Implements Professional Development). The endogenous variables associated with Problem Solving include C3Action (Action), C3ActDisp (Action Disposition), C3KTS2 (KTS 2 – Designs and



Plans Instruction), C3KTS3 (KTS 3 – Creates and Maintains Learning Climate), C3KTS4 (KTS 4 – Implements and Manages Instruction), and C3KTS6 (KTS 6 – Demonstrates the Implementation of Technology). The endogenous variables associated with Professional Leadership include C3Advocacy (Advocacy), C3AdvDisp (Advocacy Disposition), C3UofL11 (UofL Diversity Standard), C3KTS5 (KTS 5 – Assesses and Communicates Learning Results), C3KTS8 (KTS 8 – Collaborates with Colleagues/Parents/Others), and C3KTS10 (KTS 10 – Provides Leadership within School/Community/Profession).

Table 26

*Standardized Regression Weights (Loadings) for Final Model*

|            | <b>Path</b>                 | <b>Estimate</b> |
|------------|-----------------------------|-----------------|
| C3Inquiry  | <--- CriticalThinking       | .715            |
| C3InqDisp  | <--- CriticalThinking       | .733            |
| C3KTS1     | <--- CriticalThinking       | .719            |
| C3KTS7     | <--- CriticalThinking       | .672            |
| C3KTS9     | <--- CriticalThinking       | .478            |
| C3Action   | <--- ProblemSolving         | .718            |
| C3ActDisp  | <--- ProblemSolving         | .700            |
| C3KTS2     | <--- ProblemSolving         | .723            |
| C3KTS3     | <--- ProblemSolving         | .716            |
| C3KTS4     | <--- ProblemSolving         | .722            |
| C3KTS6     | <--- ProblemSolving         | .599            |
| C3Advocacy | <--- ProfessionalLeadership | .692            |
| C3AdvDisp  | <--- ProfessionalLeadership | .715            |
| C3UofL11   | <--- ProfessionalLeadership | .763            |
| C3KTS5     | <--- ProfessionalLeadership | .622            |
| C3KTS8     | <--- ProfessionalLeadership | .540            |
| C3KTS10    | <--- ProfessionalLeadership | .455            |

Table 27

*Squared Multiple Correlations for Final CF Model*

| <b>Item</b> | <b>Estimate</b> |
|-------------|-----------------|
| C3KTS10     | .207            |
| C3KTS8      | .291            |
| C3KTS5      | .387            |
| C3UofL11    | .582            |
| C3AdvDisp   | .511            |
| C3Advocacy  | .479            |
| C3KTS6      | .358            |
| C3KTS4      | .521            |
| C3KTS3      | .513            |
| C3KTS2      | .523            |
| C3ActDisp   | .491            |
| C3Action    | .516            |
| C3KTS9      | .229            |
| C3KTS7      | .451            |
| C3KTS1      | .517            |
| C3InqDisp   | .537            |
| C3Inquiry   | .511            |

The standardized regression weights (Table 26) in the final model show adequate loadings for all items (Stevens, 2009). The squared multiple correlations (Table 27) reveal that critical thinking explained 51.1% of the variance in Inquiry rating, 53.7% of the variance in Inquiry Disposition rating, 51.7% of the variance in KTS 1 (Demonstrates Knowledge of Content), 45.1% of the variance in KTS 7 (Reflects on and Evaluates Teaching and Learning), and 22.9% of the variance in KTS 9 (Evaluates Teaching and Implements Professional Development). Problem solving explained 51.6% of the variance in Action, 49.1% of the variance in Action Disposition, 52.3% of the variance in KTS 2 (Designs and Plans Instruction), 51.3% of the variance in KTS 3 (Creates and Maintains Learning Climate), 52.1% of the variance in KTS 4 (Implements and Manages

Instruction), and 35.8% of the variance in KTS 6 (Demonstrates the Implementation of Technology). Professional leadership explained 47.9% of the variance in Advocacy, 51.1% of the variance in Advocacy Disposition, 58.2% of the variance in UofL Diversity Standard 11 (Understands the Complex Lives of Students and Adults in Schools and Society), 38.7% of the variance in KTS 5 (Assesses and Communicates Learning Results), 29.1% of the variance in KTS 8 (Collaborates with Colleagues/Parents/Others), and 20.7% of the variance in KTS 10 (Provides Leadership within School/Community/Profession)

### **Summary of Results**

Results of both qualitative and quantitative analyses revealed strengths and weaknesses of the CEHD's assessment system. The qualitative analysis revealed that the current Student Learning Outcome (SLO) action plans do not include explicit references to data sources; however, those references are assumed to be implied through the statements provided by faculty that align with items assessed by the 10 Unit Key Assessments, Hallmark Assessment Tasks (HATs), and indirect survey data (QMS). The statements identified in the qualitative case study analysis demonstrate that faculty are using multiple measures of student learning and applying the findings from those measures to drive continuous improvement, which is a clear strength of the system. The SLO action plans and reflections on previous year's plans for improvement demonstrate that faculty are using data to drive improvement of curriculum, faculty development, candidates, assessments, and field/clinical experiences. It may be concluded that the CEHD's assessment system includes processes that engage faculty in the application of multiple assessment measures of teacher candidate learning to drive improvement in the

identified areas. With that conclusion, there is opportunity for improvement of the system. The analysis conducted for this study will support specific recommendations for modifications to the CEHD's assessment system. Recommendations based on findings are provided in Chapter V.

The quantitative analysis, using Hierarchical Linear Modeling, demonstrated that there was some significant variability in outcomes across programs. However, the proportion of variance that can be attributed to between-program variance was relatively low for the outcomes measured in the 10 Unit Key Assessments. The Impact assessment, which is based on the teacher candidates' reflection on his/her impact on P-12 student learning had the smallest ICC of .0246, which indicated only 2.46% of the variance in the Impact ratings could be attributed to between-program variability. This indicates excellent reliability across programs and assessors for the Impact assessment. Cycle 1 assessment, which reflects the overall observation score for the first cycle of student teaching had the highest ICC of .2347, which indicated that 23.47% of the variance in Cycle 1 ratings could be attributed to between-program variability. While this may have been a large enough ICC value to warrant some concern, there may be other potential causes for this between-program variability as the ICCs decrease from Cycle 1 to Cycle 2, Cycle 2 to Cycle 3, and Cycle 3 to Cycle 4. For Cycle 4 of student teaching, there was no significant variability ( $p=.141$ ) in outcomes by program. Additional contextual analysis revealed that the significant between-program variability could partly be attributed to degree type (B.S. or MAT) or program type (Elementary, Middle/Secondary, or Special Education) for some of the outcomes. There were no significant concerns with reliability of instruments as presented in this study.

The quantitative analysis using Structural Equation Modeling (SEM) revealed acceptable model fit for the proposed model for assessing the latent constructs of critical thinking, problem solving, and professional leadership. The Confirmatory Factor Analysis (CFA) conducted on the proposed model revealed some potential concerns with discriminant validity, as the three latent constructs all had correlations greater than the recommended .85 (Kline, 2011). While discriminant validity may partially be related to the connections between the three constructs as defined by the CEHD's conceptual framework, instrumentation issues, such as the use of a 3-point scale, may also impact the discriminant validity. This issue will be addressed in chapter V, and recommendations for next steps in improving the measurement of these three latent constructs are provided.

## CHAPTER V

### DISCUSSION

This study was designed to serve as a formative metaevaluation to identify strengths and weaknesses in the College of Education and Human Development's (CEHD) assessment system with the purpose of driving improvement of the assessment system and support building a culture of assessment within the CEHD. The primary focus of the metaevaluation was to evaluate if the CEHD assessment system provides valid and reliable data related to CEHD initial certification teacher education candidate knowledge, skills, and dispositions and to support faculty in ensuring continuous improvement and data driven decision-making, in alignment with the Council for Accreditation of Educator Preparation (CAEP) standards (2013) and the Joint Committee on Standards for Educational Evaluation (JCSEE) Program Evaluation Standards (2011). Beyond the contextual impact of this metaevaluation, findings from this study will inform the assessment and evaluation community in practices for closing the loop in assessment practices.

In alignment with Stufflebeams's (2001) structure for metaevaluation, this study presents major findings from the qualitative and quantitative analysis presented in Chapter IV and judges adherence to the CAEP standards (2013) and the JCSEE Program Evaluation Standards (2011). CAEP standards (2013) that most closely align with the

assessment system and metaevaluation of the assessment system are provided in Table 3. JCSEE Program Evaluation Standards (2011) from the Metaevaluation Checklist (Stufflebeam, 1999) that most closely align with the intent of this study are provided in Table 10. Both standard sets are incorporated into the discussion of findings for each of the four research questions.

### **Assessments Driving Program Improvement**

The first research question focused on what type of teacher candidate performance assessments best support continuous improvement in teacher education programs. This study looked specifically at which assessments of student learning were identified in the action plans of Student Learning Outcome (SLO) reports to support changes. In alignment with CAEP Standards (2013) provided in Table 3, this research question was focused on the use of “multiple measures to monitor candidate progress” and “completer achievements”.

Upon initial review of Student Learning Outcome (SLO) report action plans and reflection on last year’s plan for improvement, a weakness of the system was revealed. The action plan is a space for faculty to reflect on data and articulate plans based on the data. While it is implied that the data discussed in the findings sections of the SLO report are the basis for the action plans, the action plans are missing the detailed connection to data. For instance, only one of nine initial certification programs explicitly stated that the plan was based on the findings from analysis of the 10 Unit Key Assessments, which are the primary data sources used in the SLO reports. To demonstrate a greater connection between the data and the planned action, ideally, faculty would cite the specific data source for the planned action. The absence of the explicit connection between findings

and action plans can potentially be attributed to the SLO process and the template that is provided to faculty. To address this weakness, it is recommended that the CEHD assessment coordinator collaborate with the CEHD accountability coordinator, Office of Academic Affairs and Unit Effectiveness (AAUE) team, and the Executive Director in the Office of Institutional Effectiveness to draft revised, explicit prompts for the action plans to more explicitly request that actions be grounded in selected findings from the data discussed in the SLO reports. This recommendation is focused on supporting faculty in making meaningful reflections using data to drive plans, actions, and continuous, program improvements that are justified as part of data-driven decision-making processes and building a culture of assessment among faculty.

Looking beyond the absence of explicit connections to data in the SLO action plans, the results chapter demonstrates clear strengths in faculty application of data from the assessment system through the identified references to outcomes assessed in the 10 Unit Key Assessments, Hallmark Assessment Tasks (HATs), and student opinion survey data (QMS). These references demonstrate that the assessment system and the engagement of faculty as key stakeholders in the use of the system was focused on multiple performance assessment measures of teacher candidate learning. In addition to these three areas, there were also references in the SLO action plans to the University's Quality Enhancement Plan (QEP), Ideas to Action (i2a) Critical Thinking initiative. The critical thinking initiative was another force identified as driving program improvement through the SLO action plans and reflection about the program's previous year's action plans. This strength demonstrates faculty connections to university initiatives and, ultimately, the mission and vision of the university and college.



## **Impact of the Assessment System**

The second research question was focused on what type of change occurred as a result of application of the data and findings from the CEHD's assessment system. CAEP Standards 5.1 through 5.5 (2013) are closely aligned with this research question. The standards focus on the collection of multiple performance assessment measures of candidate learning, tracking of results, application of results to support improvement of program elements and processes, and shared summaries of results.

The analysis for this question was part of the qualitative case study analysis completed to address the first two research questions. The initial analysis of the SLO action plans and reflection on the program's previous year's plan for improvement (action plan) revealed five areas of articulated change. Those areas included improvement of curriculum, faculty development, candidate performance, assessments, and field/clinical experiences. This demonstrates a true strength of the assessment system to drive change in candidate performance through direct change that impacts candidate performance or through modifications to curriculum, field and clinical experiences, faculty development, or assessments used to measure candidate performance. The SLO process, in general, also lends to the use of results, summaries of results, and the application of findings from teacher candidate performance assessments.

This component of the SLO process could potentially be further strengthened by more extensive follow-up on the action plans to ensure that faculty are implementing the changes that they have outlined in their annual, SLO action plans. As mentioned in the results chapter there were several instances in which faculty did not respond to the prompt for reflection on last year's plan for improvement, which serves as a space to

document how faculty would go about addressing the action plan items. While the action plan provides a space for faculty to create their plan and essentially close the loop in the application of data, the reflection of the action plan or previous year's plan for improvement provides an extra layer of accountability for implementation of those plans. It may also be a space to document how plans may have been modified based on additional data, policy reform, and/or state regulation changes from the past academic year. As a result of these findings, it is further suggested that more detailed prompts in the SLO template reflection section, in addition to the action plan section, may support more meaningful reflective statements.

### **Reliability of the Assessment System**

The third research question was focused on the reliability of instruments used to measure teacher candidate knowledge, skills, and dispositions in the CEHD's assessment system. CAEP standards 3.3, 4.3, and 4.4 (2013) all reference the requirement of reliability of instrumentation in the unit's assessment system. Hierarchical Linear Modeling (HLM) was used to address reliability of measures across programs. The analysis was used to determine variability of outcomes based on teacher candidates nested in programs. The intraclass correlation coefficient (ICC) value was calculated for each outcome in the assessment system to determine the proportion of variance in the outcome that could be attributed to between program variability. The majority of ICC values were low, indicating that measures were consistent and reliable across programs and, ultimately, assessors (Raudenbush et al., 2000), as faculty often advise, instruct, and assess teacher candidates within specific programs aligned with their own training and

research areas. The only teacher candidate performance assessments of concern were the observation scores from the first two cycles of the student teaching semester.

While there could be many reasons for the variability during those initial cycles of student teaching (such as prior knowledge, skills, and dispositions), to support reliability of measures, the CEHD would benefit from sessions for collaborative review of rubrics and practice application of instruments with all faculty and university supervisors who supervise student teaching and who complete the observation forms. University supervisors currently participate in supervisor training sessions prior to the beginning of the student teaching semester, and it is recommended that opportunities for rubric dissection and discussion of scoring processes beyond what is currently required be incorporated to address reliability. Further, it is suggested that these findings be discussed collaboratively with the Office of Educator Development and Clinical Practice (OEDC), department chairs, program coordinators, the College Educator Preparation Committee (CEPC), and members of the AAUE team to ensure systematic implementation of training and support to field evaluators.

### **Validity of the Assessment System**

The fourth and final research question was focused on the validity of instruments in the CEHD assessment system. CAEP Standards 3.3, 4.2, 4.3, 4.4, and 5.2 (2013) all address the requirement for valid data in the unit's assessment system. Confirmatory Factor Analysis (CFA) was conducted to review construct validity of assessments as they relate to the CEHD's conceptual framework. A clear strength of the items measured within the CEHD system was the relatively large loadings for items intended to measure the latent constructs of critical thinking, problem solving, and professional leadership

based on the design of the assessment system. This demonstrates that the conceptual framework constructs are truly embedded within the 10 Unit Key Assessment and, ultimately, the CEHD's assessment system. Construct validity demonstrates the connection of the CEHD's assessment system to the mission and vision of the college and the university as a whole.

While the loadings demonstrated promising construct validity and the proposed conceptual model based on the design of the CEHD assessment system instrumentation were adequate, there was one weakness related to discriminant validity. As discussed in the results chapter, the latent constructs of critical thinking, problem solving, and professional leadership were all highly correlated ( $>.85$ , Kline, 2011). These correlations may be partly a result of the relationship between the constructs as defined by the CEHD's conceptual framework; however, the use of the 3-point scale to assess candidate knowledge, skills, and dispositions may also be contributing to the issue of discriminant validity.

Prior to this metaevaluation, AAUE team members and CEHD faculty have engaged in multiple discussions about transitioning assessment rubrics to four or five point scales for assessing the 10 Unit Key Assessments and Hallmark Assessment Tasks (HATs) to provide more discriminant results regarding candidate performance assessment. There is also evidence of this movement to move this transition of instrumentation forward as faculty have engaged in pilots of new instruments. One program recently piloted a 4-point rubric during the student teaching capstone semester. The faculty completed the original 3-point rubric and the new 4-point rubric. After completion of the pilot, faculty reviewed the data from both rubrics during a program

faculty meeting. The assessment coordinator was present to facilitate the discussion, which resulted in approval of the transition to the 4-point rubric going forward, with continued development of rubrics and training for all assessors. The faculty expressed that the 4-point rubric provided them with more discriminant data surrounding teacher candidate performance assessment to support data-based decision making around program improvement. This pilot demonstrates that the suggested revision of assessment rubrics will be in alignment with stakeholder goals. Further analysis of the pilot rubric is recommended to ensure that the 4-point scale will address concerns with discriminant validity.

To support the transition to 4-point scales for all rubrics used in the CEHD's assessment system, it is suggested that the AAUE team develop protocols to support faculty in engaging in the rubric revision process. It is recommended that the CEHD consider hiring an external leader in the assessment field to provide training on rubric development to provide faculty with best practices and approaches to revising program rubrics. Based on the assessment system's heavy emphasis on teacher candidate performance assessment data from the 10 Unit Key Assessments, it is suggested that these assessments be transitioned first. Next, faculty should continue to modify Hallmark Assessment Tasks (HATs) and other course assessments. The AAUE team can provide protocols and timelines for revision and implementation to streamline the transition process. In addition to training from an external leader in rubric development, it is further recommended that faculty training sessions to review revised rubrics and ensure inter-rater reliability be implemented. These sessions should be facilitated by program faculty

who have experience within this area and have the knowledge of programmatic contexts in which the rubrics are applied.

One important component of the rubric development will be differentiating between formative and summative assessment in the program. Faculty involved in the pilot of the 4-point rubric had valuable discussion about the balance between using rubrics to guide students towards Target (highest level), versus generating meaningful data that identifies strengths and weaknesses in candidate performance. The use of rubrics to guide student performance is important and represents formative assessment; however, if the intent of the rubric is to guide students to Target, then it may limit the applicability of assessment results towards driving program improvement. This thoughtful discussion with faculty further demonstrates the engagement and ownership of the CEHD assessment system by faculty as the primary stakeholders.

### **Alignment to CAEP Standards**

This study was closely aligned with the CAEP Standards (2013) to evaluate the CEHD's assessment system in alignment with external accountability from accrediting agencies. The following narrative is a review of each of the CAEP standards provided in Table 3 with discussion of findings from this study.

***CAEP Standard 3.3.*** Standard 3.3 was directly addressed by this study. The elements and assessments within the CEHD's assessment system, outlined in chapter I, demonstrate that the CEHD has established measures of candidate performance beyond academic ability and monitors candidate attributes and dispositions. This study addressed the reliability and validity of the instruments used to measure knowledge and dispositions within the assessment system. The findings identified that the instruments had adequate

reliability and construct validity; however, instruments should be revised to address discriminant validity concerns.

**CAEP Standard 3.4.** Standard 3.4 focuses on the transitional tracking of teacher candidates and their ability to demonstrate developing knowledge, skills, and dispositions through evidence. The embedding of the 10 Unit Key Assessments, which focus on knowledge, skills, and dispositions, across the phases of admissions, mid-program, and exit, demonstrates that programs are tracking candidate development with evidence.

**CAEP Standard 4.1.** Standard 4.1 is focused on connections to P-12 student learning. Unit Key Assessment 6 is an assessment of Impact on P-12 Student Learning. This assessment looks specifically at the candidate reflection on P-12 student learning from their student teaching semester. In addition to assessments included in this study, it is recommended that the AAUE team engage in regular discussion of the Kentucky Department of Education's implementation of the new Professional Growth and Effectiveness System (PGES, 2013), which will eventually connect P-12 student performance to teacher preparation programs. The PGES will have the potential to provide greater understanding, through quantitative measures, the impact of teacher preparation and subsequent teacher efficacy at each higher education institution in the state. In addition to the PGES, further emphasis and review of Kentucky Teaching, Empowering, Leading, and Learning (TELL) survey data should be incorporated into the CEHD assessment system (Kentucky TELL, 2013). The TELL survey was administered in 2013 by the New Teacher Center (NTC) and provides feedback from P-12 educators, by county, on topics such as managing student conduct, time management, and new

teacher support which has the potential to support continuous improvement of teacher preparation programs.

**CAEP Standard 4.2.** Standard 4.2 requires that programs demonstrate that candidates are effectively applying the knowledge, skills, and dispositions outlined by the program in the student teaching experience. This standard aligns with the four formal observations that are captured during student teaching and are included in the 10 Unit Key Assessments. Faculty document and analyze these data in the annual Student Learning Outcome (SLO) report and further require an acceptable or higher level of teacher candidate performance to ensure adequate learning. Further, the instruments used for observation are aligned with the Kentucky Teacher Standards (KTS) which are also embedded throughout multiple course assessments, Hallmark Assessment Tasks (HATs), and the 10 Unit Key Assessments.

**CAEP Standard 4.3.** Standard 4.3 is focused on ensuring the use of valid and reliable data as well as data that support employer satisfaction with teacher candidates' training. This study focused on the reliability and validity of instruments within the CEHD's assessment system. As previously discussed, this study encourages revision of assessment rubrics to address potential discriminant validity concerns. Holistically, the review of the 10 Unit Key Assessments revealed reliable measures, with adequate construct validity in respect to the CEHD's conceptual framework. While employment milestones and employer satisfaction were not discussed previously in this study, the CEHD has provided a data dashboard with additional data sources for faculty review in addition to teacher candidate performance assessment data from the 10 Unit Key Assessment process. The data dashboard is linked to the Kentucky Education



Professional Standards Board's (EPSB) data dashboard with employment statistics and data from the CEHD's employer survey of CEHD graduates. The AAUE curriculum coordinator also compiled a summary report of Kentucky New Teacher Survey data (administered by EPSB), TELL survey data (KDE), and the CEHD's employer survey and engaged program faculty in a review of the major findings from these surveys during a faculty retreat in the Fall of 2013. The increased emphasis on educator preparation program (EPP) completer data will continue to require that these data be incorporated into the CEHD's assessment system and that faculty regularly discuss these data as part of the culture of assessment.

***CAEP Standard 4.4.*** Standard 4.4 is focused on ensuring the use of valid and reliable data, as well as data that support program completer satisfaction in their training. This study focused on the reliability and validity of instruments within the CEHD's assessment system. As previously discussed, findings from this study suggest the need for revision of rubrics to address potential discriminant validity concerns. Holistically, the review of the 10 Unit Key Assessments revealed reliable measures, with adequate construct validity in respect to the CEHD's conceptual framework. Further, data are needed to address completer satisfaction with their training as these data were not included in this study. There are several data sources incorporated in the CEHD's assessment system that address this component. The QMS survey data, which is a student satisfaction survey, was well represented in the SLO reports as an indirect assessment to support continuous improvement. It is suggested that CEHD programs also include references to New Teacher Survey data (EPSB) collected at the state level, Kentucky TELL survey data (KDE), as well as the CEHD employer survey data, within the Student

Learning Outcomes (SLO) report, to make these connections more explicit in addressing this standard.

**CAEP Standard 5.1.** Standard 5.1 requires that the educator preparation provider (EPP) have a quality assurance system in which multiple measures of candidate progress and achievements are collected. The CEHD's comprehensive assessment system aligns with this expectation and ensures that the faculty are engaged in this use of this system. The 10 Unit Key Assessments demonstrate the multiple measures that capture candidate performance assessment, progress, and achievements.

**CAEP Standard 5.2.** Standard 5.2 focuses on the educator preparation provider (EPP) having an assessment system that provides "relevant, verifiable, representative, cumulative and actionable measures, and produces empirical evidence"(CAEP, 2013). This standard is clearly a strength area of the CEHD due to the extensive data captured within the assessment system. The measures within the assessment system are aligned with state and national standards and the unit's conceptual framework and are used to drive continuous improvement of teacher preparation programs as shown by the analysis of the SLO reports.

**CAEP Standard 5.3.** Standard 5.3 describes the CEHD's assessment system as outlined in this study. The standard states that programs should "regularly and systematically assess performance against program goals and relevant standards", which is demonstrated by the connections of the CEHD's assessment system to the CEHD's conceptual framework and the Kentucky Teacher Standards (KTS). This standard further focuses on the use of "results to improve program elements and processes," which is demonstrated through application of data in the SLO reports. The CEHD should continue

to regularly incorporate internal teacher candidate performance assessment data as well as state-level data (New Teacher Survey and TELL survey) into department and program meetings, and curriculum planning to support a culture of assessment and continuous improvement among CEHD faculty.

***CAEP Standard 5.4.*** Standard 5.4 is focused on P-12 student growth data.

Although not a direct focus of this study it is important to note that these data will soon be available through the Kentucky Education Professional Standards Board's (EPSB) implementation of the Professional Growth and Effectiveness System (PGES, 2013). Further, collaboration with state agencies and other education preparation programs across the state is encouraged to address this standard.

***CAEP Standard 5.5.*** Standard 5.5 focuses on involvement of key stakeholders, such as school and community employers, in program evaluation and improvement.

While the involvement of external stakeholders was not addressed in this study, these stakeholders are involved at many levels of the assessment system as demonstrated by the CEHD Assessment System Feedback Loop (Figure 1). The University Educator Preparation Committee (UEPC) has multiple representatives from surrounding school districts, as well as leaders for community engagement, at the university. The UEPC also provides a mechanism to engage practicing P-12 professionals in driving continuous improvement of educator preparation programs. This involvement ensures that the system is responsive to these critical stakeholders.

This review of the CAEP Standards (2013) helps to identify some potential next steps in the review of the CEHD's assessment system and educator preparation programs. There were components of the standards that were not a direct focus of this study.

However, there is evidence that these items are being addressed within the CEHD current operations and assessment system. In alignment with the standards, and the findings provided in Chapter IV, the CEHD has demonstrated a comprehensive evaluation system that guides program improvement and supports a culture of assessment. Continued improvement of assessment instruments, reporting templates, and faculty development in assessment and creating a culture of assessment, will ensure that the unit continues to address these standards and evolving accountability in the field of teacher preparation.

### **Adherence to JCSEE Program Evaluation Standards**

The Joint Committee on Standards for Educational Evaluation (JCSEE), Program Evaluation Standards have been recommended as an instrument to judge quality in evaluation and metaevaluation practices (Stufflebeam, 1999). The following narrative provides a discussion of adherence to the standards provided in Table 10. These standards were selected prior to conducting this study as standards aligned with the purpose of the assessment system in alignment with CAEP standards. The standards are organized in the areas of propriety, accuracy, feasibility, and utility.

***Propriety (1) - Service Orientation.*** P1 is focused on whether or not the CEHD's assessment system is serving primary stakeholders and supporting their ability to identify strengths and weaknesses within their programs to drive improvement. As demonstrated in the Student Learning Outcome (SLO) reports, faculty have been provided structures for the review of candidate performance data that can then be used to determine action plans for the coming academic year. Analyses revealed that the SLO process supports stakeholders in engaging in the use of data and identifying ways to improve programs.

***Propriety (4) - Human Interaction.*** P4 is aligned with ensuring that stakeholder engagement is seamless and strategic to ensure minimal disruption to other professional activities. The use of the Student Learning Outcome (SLO) report as a mechanism to drive the application of data for program improvement demonstrates a strategic incorporation of already present structures into the CEHD's assessment system. Since the SLO report is a university-wide requirement, it supports sustainability of the process for the CEHD's assessment system. Embedding the CEHD's accountability structures into the university accountability structures and processes demonstrates further alignment with university goals in the work of the CEHD.

***Accuracy (5) - Valid Information.*** A5 "encourages the use of multiple measures" in addition to documentation of data collection processes. As demonstrated by the 10 Unit Key Assessments and Hallmark Assessment Tasks (HATs), there are multiple measures of teacher candidate learning within the assessment system. Further, the CARDS transitional phases of admissions, mid-program, and exit demonstrate the explicit and transitional documentation of data collection processes. This study also looked more specifically at the construct validity of instruments that were developed in alignment with the CEHD's conceptual framework constructs of Inquiry, Action, and Advocacy. While construct validity was considered acceptable for those items, further revision of the system to address discriminant validity is recommended based on the findings of this study.

***Accuracy (6) - Reliable Information.*** A6 "encourages awareness and justification of reliability with a review of factors that influence reliability." This metaevaluation looked specifically at reliability of instruments across programs. The 10 Unit Key

Assessments were designed to be used across multiple, diverse programs that focus on teacher training in multiple grade levels and content areas. The awareness of the differences between these programs led to this study's investigation of reliability. In addition to reliability of the 10 Unit Key Assessment addressed in this study, the electronic system used to capture data for Hallmark Assessment Tasks (HATs) automatically generates inter-rater summary data for rubrics that are used across raters.

*Accuracy (7) – Systematic Information.* A7 is concerned with ensuring systematic collection and storage of evaluation data. The CEHD's assessment system was developed with transitional phases, or milestones, that ensure systematic tracking of candidate progress. Further, the integration of the assessment system into existing university data systems (PeopleSoft©), as well as reporting structures such as SLOs, demonstrates a systematic approach to information collection, storage, and application. The staff in the Office of Academic Affairs and Unit Effectiveness (AAUE) in the CEHD further support systematic oversight and verification of data to ensure accurate and complete records.

*Accuracy (10) – Justified Conclusions.* A10 focuses on the assurance that decisions are being based on evidence aligned with evaluation questions. The SLO process lends well to ensuring justified conclusions as it is a process that relies closely on the 10 Unit Key Assessments which are intended to address important outcomes in the field of teacher preparation. Further, the SLO process requires faculty to reflect on the findings from the 10 Unit Key Assessments and to develop action plans based on those reflections to drive program improvement.

***Accuracy (11) – Impartial Reporting.*** A11 addresses the importance of engaging multiple stakeholders in the review of assessment findings. Faculty work groups regularly engage in the reflection of data for the Student Learning Outcomes (SLO) report process. Data that are reviewed focus on assessments from a variety of faculty and field supervisors with varying training and experience. Further, involvement of external stakeholders is incorporated into the assessment system through university educator preparation programs as well as P-12 partnerships. There is substantial evidence of impartial reporting in the CEHD’s assessment system.

***Feasibility (1) – Practical Procedures.*** F1 is focused on ensuring that the assessment system does not place a burden on faculty or have unrealistic processes. It further encourages the embedding of evaluation procedures into already present structures and processes. The CEHD’s integration of the university’s SLO process into the assessment system to ensure that faculty have a mechanism to use data and apply findings towards driving continuous improvement demonstrates that the CEHD has been mindful of faculty time and energy to ensure that faculty are engaged in the process, but without substantial burden. The use of practical procedures in the CEHD assessment system is clearly a strength because of this use of already present processes.

***Utility (7) – Evaluation Impact.*** U7 states, “involves stakeholders throughout and encourages use of findings” (2011). This standard demonstrates the importance of ensuring that the key stakeholders can own and use the evaluation process to meet their goals. In the case of the CEHD’s assessment system, program faculty are the primary stakeholders. Findings from this study indicate that faculty are engaged in the development and application of assessments, analysis of data from their assessments, and

the application of findings from those assessments as demonstrated through the Student Learning Outcome (SLO) reports. Based on the qualitative case study analysis of SLO action plans, faculty are applying findings from their assessments to identify plans for improvement in regards to curriculum, assessments, faculty development, field and clinical experiences, and candidate performance. This clearly demonstrates that the assessment system has an impact on the quality of programs in the CEHD.

This review of the JCSEE program evaluation standards (2011) demonstrates that the CEHD faculty, staff, and administrators have developed a system for assessment that supports continuous program improvement. The system is making an impact on program quality and ensures data-based decision making that is practical, yet strategic. The oversight of the system by AAUE staff ensures that data collection and reporting are supported and verified to support faculty application of the system. As with any system, there is always room for improvement, and this metaevaluation is the next step in addressing ways to improve the system for continued unit and program improvement.

### **Recommendations for Next Steps**

Based on the results in Chapter IV, the review of alignment with CAEP standards (2013), and the adherence to the JCSEE Program Evaluation Standards (2011) as outlined above, there are several strengths and weaknesses of the CEHD's assessment system. Strengths include the systematic and strategic processes for the collection and application of data, reliable instrumentation, and instrumentation that is aligned with the CEHD's conceptual framework. The primary weakness of the CEHD's assessment system is the use of the 3-point scale for rubrics used to measure teacher candidate performance assessments. Additional areas for improvement include more explicit connections to data



in the plans for improvement, greater follow-up on implementation of plans for improvement, and faculty development for designing and implementing high quality and discriminant assessment rubrics.

This study was designed to be a formative metaevaluation, with the intent to inform next steps through identification of strengths and weaknesses. This study was not intended to be a judgment of quality or summative metaevaluation. Based on the qualitative and quantitative analyses conducted in this study, there are several recommendations for next steps in improving the CEHD's assessment system.

The following next steps have been guided by input from the CEHD's Vice Dean, the CEHD's accountability coordinator, and the university's Executive Director of Institutional Effectiveness, as well as other members of the Office of Academic Affairs and Unit Effectiveness (AAUE) in the CEHD. The Vice Dean brings extensive knowledge of accreditation and accountability expectations in the field of educator preparation as well as years of experience as a teacher education program faculty member. The accountability coordinator has and continues to provide valuable input into the development, implementation, and continued advancement of systematic processes aligned with accountability expectations. The Executive Director of Institutional Effectiveness oversees the university's Student Learning Outcomes (SLO) report process, served as an external stakeholder who provided valuable input into this study, and provided connections to university perspectives about strengths and weaknesses of the SLO process and template.

### **Revision of prompts in Action Plans**

As revealed in the qualitative case study analysis of action plans, the Student Learning Outcome (SLO) reports are missing explicit connections between the findings from the 10 Unit Key Assessments and the action plans. It is recommended that additional prompts be added to the Action Plans to support making these connections. The addition of prompts that explicitly request the data source and rationale for change in the action plans will help to demonstrate more explicit connections to teacher candidate performance assessment data to show evidence of thoughtful data-driven decision-making to guide program improvement.

In addition to the annual SLO report completed in the Fall, the CEHD has implemented a mid-year report, which serves as an opportunity for faculty to revisit their action plans and document potential changes, as well as document their progress in addressing elements included in the action plan. The assessment coordinator and accountability coordinator have already worked together to add the table shown in Figure 9. This table is designed to move faculty to work toward making connections between the assessment system data and improvement decisions that they make within their programs. While evidence of data-based decision-making may seem explicit from the view of the assessment coordinator and others in working with faculty, it is recommended that the AAUE team work more closely with faculty to create processes that support more explicit connections to multiple and high quality data and further generate awareness of connections to outcomes for both internal and external accountability purposes.

*Mid Year SLO Report Connection Prompts*

**In the space below, document 3 or more examples of how your program has used student outcome data to support a program, course, assessment, or other change in the past academic year. Identify the specific data source that informed the change and the change that was made or is planned.**

| <b>Identify the data that informed the change (such as, SLO, survey, evaluation, QMS, etc.)</b> | <b>Describe the change that was made (include specific information, such as the name of the program, course, assignment, and exactly what was done, etc.)</b> | <b>Semester and year</b> |
|---|---|--------------------------|
| 1.  |   |                          |
| 2.  |   |                          |
| 3.  |   |                          |
| [add more rows as needed]   |   |                          |

Figure 9. Mid Year SLO Report Revision. The prompt will be added to the Spring 2014 Mid Year SLO Report to facilitate faculty identifying and documenting explicit connections between data and actions.

It is also recommended that the AAUE staff continue to work closely with the CEHD Vice Dean and the Executive Director of Institutional Effectiveness for the university to review the impact of adding these additional prompts to ensure that they positively impact the application of data from the CEHD’s assessment system. The required SLO training sessions provided to faculty by the AAUE staff are one recommended mechanism to increase awareness of the importance of these components and the need for explicit connections between data and action plans for program improvement.

**Increased Emphasis on the Reflection on Previous Year’s Plan for Improvement**

It is proposed that in addition to prompts in the SLO action plans, there should also be greater emphasis on the follow-through of action plans. The addition of Figure 9

to the mid year report will help to address this follow-through. It was identified in the qualitative case study analysis that several programs did not complete the reflection component on the previous year's plan for improvement in the SLO report template. It is recommended that greater emphasis be placed on this section during the SLO required training sessions provided every fall to faculty by the AAUE staff.

While the reflection on the previous year's plan for improvement is not a section required at the university level, it is recommended that the AAUE staff add measures to hold CEHD faculty accountable to complete this section to ensure that there is evidence of data-driven decision-making and the impact of program changes implemented based on data. Additionally, it is recommended that the AAUE staff collaboratively develop new prompts for the reflection section to ensure that this element encourages documentation that shows evidence of continuous improvement of programs as a result of the unit's comprehensive assessment system.

### **Revision of format for Student Learning Outcome Reports**

In addition to adding prompts to the SLO action plans and the mid year SLO report, it is suggested that the AAUE staff review the current SLO template format in LiveText©. The electronic template that is used for SLOs has each SLO broken out into a separate table with a column for the program goal, the student learning outcome, the measures and intended targets, and then the findings from data for the identified measures. The template leads from left to right to address these items. The action plan is then located at the bottom of the template, after all of the tables addressing the goals, outcomes, measures and targets, and findings.

In a discussion with the Executive Director of Institutional Effectiveness, she agreed that the horizontal progression from program goal to findings can impact the flow of the work into the action plan. She and the researcher discussed further revision of the SLO template that is shared as a Microsoft Word document to other units that do not use LiveText©. The Office of Institutional Effectiveness revised the template so that faculty would write the progression from program goal, student learning outcome, measures and targets, and findings in columns (top to bottom), which then gives a visual progression towards the action plan at the bottom of the template. Based on the qualitative case study analysis and advisement from the Executive Director of Institutional Effectiveness, it is recommended that the AAUE team continue to have discussions about how to modify the current LiveText© template to best guide CEHD faculty in the process and ensure connections between all of the components of the SLO template.

### **Revision to Incorporation of Feedback in SLO process**

The results of the qualitative case study analysis were discussed with the Executive Director of Institutional Effectiveness. One of her recommendations was to revisit the CEHD's current mechanisms for incorporation of feedback into the SLO process to further support quality in this process. The Office of Institutional Effectiveness provides annual feedback on SLO reports to programs. In previous academic years, the CEHD had also incorporated a round of internal peer review of SLO reports to support quality in the process.

It is recommended that AAUE staff look at potential mechanisms to incorporate feedback into the current SLO process. One suggestion would be to provide time during the CEHD SLO required training sessions for faculty to review their feedback and

discuss how they might address the comments that were made. The AAUE staff could also potentially compile sample feedback to identify some themes in the feedback and provide sample discussion about how to address the feedback during the training sessions. It is recommended that this item be added to AAUE team meeting agendas for further brainstorming and discussion.

### **Revision of Rubrics**

The Confirmatory Factor Analysis (CFA) conducted to address construct validity revealed concerns regarding discriminant validity. Supported by previous movements among CEHD faculty and AAUE staff to transition to 4-point rubrics, the AAUE should develop timely processes and procedures to move this work forward. Four-point rubrics will have the potential to provide differentiated data on candidate performance assessment that can better support program improvement efforts.

The revision of assessment rubrics will be an extensive project due to the multiple program assessments embedded within courses, field experiences, clinical practice, and other program milestones. To support this process, it is suggested the faculty work groups first revise rubrics for the 10 Unit Key Assessments and then continue to revise Hallmark Assessment Tasks (HATs), and other assessments. It is further suggested that the CEHD consider hiring an expert in the field of assessment to provide faculty development sessions on rubric development to further support this process.

An additional challenge of the transition to 4-point assessment rubrics is the potential burden on faculty to engage in this work. It is suggested that the AAUE attempt to facilitate this work during the summer semester when course loads and faculty work plans may provide for more flexibility and time to commit to this process. While it is

important to be mindful of faculty time, they are the primary stakeholders of the assessment system. Instruments require faculty extensive knowledge of content, pedagogical, and professional knowledge to ensure quality and assessments that align with the curriculum in addition to professional standards.

### **Faculty Development**

As the CEHD moves towards the development of new assessment rubrics, additional faculty development will be necessary to address reliability of instruments. It is recommended that the AAUE staff work closely with department and program chairs to utilize current meeting and work structures to engage in this work. The CEHD has faculty who are experienced in instrument development, dissection, and reliability training. It is suggested that these faculty be asked to lead sessions with other faculty groups to ensure consistent and reliable data.

As noted in the analysis of reliability in this study, it will be important for faculty to have discussions about differences among programs, among degree types, and among program types. Candidate performance may potentially vary due to content areas, undergraduate (B.S.) versus graduate (MAT), and elementary versus middle and secondary. Collective discussions of assessment rubrics across programs and also within specialization areas will further benefit faculty in the implementation of instruments.

In addition to faculty development to support revision and application of rubrics, continued development of the CEHD assessment system will be necessary to address the need for continuous improvement of the system, as well as adapt to changing accountability. In response to this need for continuous improvement and on-going faculty development in relation to the CEHD's assessment system, the CEHD Office of

Academic Affairs and Unit Effectiveness (AAUE) is encouraged to engage faculty through Appreciative Inquiry (AI) in department meetings, program meetings, and College Educator Preparation Committee Meetings (CEPC).

Appreciate Inquiry (AI) is a process for driving organizational change that builds upon existing strengths and engages stakeholders in strategic planning to drive change (Fitzpatrick, Christie, & Mark, 2009; Rothwell & Sullivan, 2005). AI consists of four stages of Discovery (review of past strengths), Dream (envisioning the future), Design (designing of structures to support the future), and Destiny (proposing and implementing actions) (Rothwell & Sullivan, 2005). Engagement of faculty in these four phases aligns with research that supports participatory evaluation techniques to improve buy-in of internal stakeholders (Payne & Miller, 2009). AI has the potential to enhance faculty engagement and moral around assessment and further support a culture of assessment among CEHD faculty, as primary stakeholders of the CEHD's assessment system.

### **Implications for Teacher Education Programs and Evaluation Field**

This study demonstrates the importance of reviewing evaluations and evaluation systems for quality. Accountability has and continues to drive organizations to engage in quality control and impacts consciousness of evaluation practices, procedures, and implementation (Mero, Guidice, & Anna; 2006). The practice of metaevaluation is a proactive approach to identifying strengths and weaknesses in an evaluation or evaluation system, which further supports identification of next steps in improving the evaluation or evaluation system (Stufflebeam, 2011). In alignment with Scriven (1969), this process of metaevaluation serves as an important function because it holds individuals and organizations accountable for the work and ensures values and purpose in the work.



This study demonstrates one approach to conducting a metaevaluation of a comprehensive outcomes-based assessment system. It has the potential to raise awareness of the need for continued review of evaluation and evaluation systems. In an era in which universities have struggled to go beyond simply developing and implementing systems for data collection and assessment (Wilkins, Young, & Sterner, 2009), metaevaluation can serve as a mechanism to address much needed next steps in improving these systems.

This study has specific implications for the field of teacher education, as institutions that prepare teacher education candidates must address accreditation standards and also continue to address policy reform in both higher education and P-12 education. Evaluation and metaevaluation are a means to address accountability from the government, the public, and accreditation agencies. The strengths identified in the CEHD's assessment system provide examples of best practices in teacher preparation program evaluation systems. The processes that have been put in place by the CEHD to ensure continuous assessment, documentation, reflection about data, and application of data for program improvement are aligned with accreditation standards (CAEP, 2013) and evaluation standards (JCSEE, 2011).

It is hopeful that this study brings awareness to metaevaluation and the value that metaevaluation can have in the assessment and evaluation field. Metaevaluation serves as a mechanism to hold evaluators accountable for evaluation quality (Fitzpatrick, Sanders, and Worthen, 2011; Stufflebeam, 2001).

### **Limitations**

The primary difference between evaluation and research is that evaluation is contextual and is not designed for the purpose of generalizability (Fitzpatrick, Sanders, &

Worthen, 2011). The components of the CEHD's assessment system reviewed in this study were designed, developed, and implemented by faculty, staff, and administrators at the University of Louisville in the context of NCATE and EPSB accreditation, Kentucky accountability structures and educational climate, University of Louisville structures, CEHD structures and capacity, P-12 partners, and many other contextual factors. Because of these contextual influences, and the unique design of assessments and instruments within the assessment system, there is limited generalizability of findings in regards to instrumentation and application of assessment findings.

### **Study Conclusions**

While accountability has been present in higher education since the founding of the first institutions (Thelin; 2011), it has been argued there have been significant shifts in accountability towards increased transparency in recent years (Mehta, 2013; Webber & Boehmer, 2008). Evaluation is an essential process to assess programs and identify strengths and weaknesses to improve quality that addresses accountability in higher education (Fitzpatrick, Sanders, & Worthen, 2011). Further, metaevaluation serves as the mechanism to ensure quality in evaluation practices (Stufflebeam, 2011). This study demonstrates the value in metaevaluation as a mechanism and essential step to addressing quality in evaluation and evaluation systems that both support and address both internal and external accountability in higher education.

Operationally, metaevaluation gathers information to determine quality of evaluations or evaluation systems in regards to utility, feasibility, propriety, and accuracy (Stufflebeam, 2001). Through the application of Stufflebeam's structure for metaevaluation (2001), this study has demonstrated the ability of metaevaluation to drive

improvement of evaluation and evaluation systems aligned with those four areas through identification of stakeholders, alignment with professional standards, and a methodologically objective review of existing information. This study was conducted as an internal evaluation because of the importance of contextual knowledge in evaluating this evaluation system (Fitzpatrick, Sanders, & Worthen, 2011). To address bias and integrity of this internal metaevaluation, dissertation committee members and other internal and external stakeholders were selected to engage in this study based on expertise in accountability, assessment, evaluation, institutional research, research methodology, and teacher education.

This study adds to the current research on best practices in evaluation and metaevaluation approaches by demonstrating the capability of metaevaluation to identify strengths and weaknesses in the CEHD's assessment system for initial teacher preparation. Understanding of the value and processes for metaevaluation has implications for the assessment and evaluation field by bringing awareness to the mechanisms that ensure quality in evaluation and evaluation systems. This study further has implications for addressing future accountability through quality evaluation practices.

Assessment systems designed to capture and report data in relation to outcomes, such as the CEHD's assessment system, have been prevalent in higher education for years (Kirchner, 2012). In alignment with evaluation practices and standards, metaevaluation is essential to ensuring quality in evaluation and evaluation systems like the CEHD's assessment system (Fitzpatrick, Sanders, and Worthen, 2011; Scriven, 1969; and Stufflebeam, 2011). This study demonstrates the ability of metaevaluation to identify

strengths and weaknesses to determine next steps in revising an evaluation system to support quality evaluation processes.

As part of the University of Louisville, College of Education and Human Development's commitment to being a premier, metropolitan research institution with a focus on institutional effectiveness, this study demonstrates the value of evaluation and metaevaluation in driving continuous improvement. Evaluation and metaevaluation are both essential functions of organization as they serve to ensure quality of programs and services, and ensure accountability (Scriven, 1969), in this case, initial teacher preparation programs. It is hopeful that this study will drive further advancement of evaluation and evaluation systems at the institutional level and, ultimately, impact quality in evaluation in higher education.

## REFERENCES

- Adelson, J. L. (2012). Examining Relationships and Effects in Gifted Education Research: An Introduction to Structural Equation Modeling. *Gifted Child Quarterly*, 56(1), 47-55. doi:10.1177/0016986211424132
- American Evaluation Association Guiding Principles for Evaluators. Retrieved from <http://www.eval.org/p/cm/ld/fid=51>
- Amobi, F. A. (2003). Reflections on the Transformation in a Teacher Educator's Teaching of Educational Philosophy. *Teacher Education Quarterly*, 30(3), 23-32.
- Arena, M., Arnaboldi, M., Azzone, G., & Carlucci, P. (2009). Developing a Performance Measurement System for University Central Administrative Services. *Higher Education Quarterly*, 63(3), 237-263. doi:10.1111/j.1468-2273.2008.00415.x
- Association of American Colleges and Universities (AACU). (2002). Greater Expectations: A New Vision for Learning as a Nation Goes to College. Retrieved from <http://greaterexpectations.org/>
- Barrett, P. (2007). Structural equation modelling: adjudging model fit. *Personality and Individual Differences*, 42, 815–824.
- Bemak, F., & Chung, R. (2005). Advocacy as a Critical Role for Urban School Counselors: Working toward Equity and Social Justice. *Professional School Counseling*, 8(3), 196.
- Bentler, P. M., & Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin*, 88, 588-606.
- Bhatt, R. and Koedel, C. (2012). Large-Scale Evaluation of Curricular Effectiveness: The Case of Elementary Mathematics in Indiana. *Educational Evaluation and Policy Analysis* 34(4), 391-412.
- Bibens, R. F. (1980). Using Inquiry Effectively. *Theory Into Practice*, 19(2), 87-92.
- Bickman, L. (1997). Evaluating evaluation: Where do we go from here?. *Evaluation Practice*. 18(1), 1-16.

- Bledsoe, K. L., & Graham, J. A. (2005). The Use of Multiple Evaluation Approaches in Program Evaluation. *American Journal Of Evaluation*, 26(3), 302-319.
- Bornmann, L. L., Mittag, S. S., & Daniel, H. D. (2006). Quality Assurance in Higher Education--Meta-Evaluation of Multi-Stage Evaluation Procedures in Germany. *Higher Education: The International Journal Of Higher Education And Educational Planning*, 52(4), 687-709.
- Braverman, M. T. (2013). Negotiating Measurement: Methodological and Interpersonal Considerations in the Choice and Interpretation of Instruments. *American Journal Of Evaluation*, 34(1), 99-114. doi:10.1177/1098214012460565
- Carlson, M. P., & Bloom, I. (2005). The Cyclic Nature of Problem Solving: An Emergent Multidimensional Problem-Solving Framework. *Educational Studies In Mathematics*, 58(1), 45-75.
- Carman, J. G. (2013). Evaluation in an Era of Accountability: Unexpected Opportunities – A Reply to Jill Chouinard. *American Journal Of Evaluation*, 34(2), 261-265. doi:10.1177/1098214013478145
- Chouinard, J. (2013). The Case for Participatory Evaluation in an Era of Accountability. *American Journal Of Evaluation*, 34(2), 237-253. doi:10.1177/1098214013478142
- Chouinard, J. (2013). The Practice of Evaluation in Public Sector Contexts: A Response. *American Journal Of Evaluation*, 34(2), 266-269. doi:10.1177/1098214013478146
- Cohen, J. (1977). *Statistical power analysis of the behavioral sciences* (rev. ed.). New York: Academic.
- Cooksy, L.J., & Caracelli, V.J. (2005). Quality, context, and use: Issues in achieving the goals of metaevaluation. *American Journal of Education*, 26(1), 31-42.
- Council for the Accreditation of Educator Preparation (CAEP). <http://caepnet.org>.
- Council for the Accreditation of Educator Preparation (CAEP) Standards (2013). <http://caepnet.org/accreditation/standards/>.
- Creswell, J.W. (2012). *Educational Research: Planning, Conducting, and Evaluating Quantitative and Qualitative Research* (4<sup>th</sup> edition). Boston, MA: Pearson Education, Inc.
- Darling-Hammond, L., & Baratz-Snowden, J. (2007). A Good Teacher in Every Classroom: Preparing the Highly Qualified Teachers Our Children Deserve. *Educational Horizons*, 85(2), 111-132.

- Dew, J. R. (2012). The Future of American Higher Education. *World Future Review*, 4(4), 7-13.
- Duck, L. (2000). The Ongoing Professional Journey. *Educational Leadership*, 57(8), 42-45.
- Eaton, J. S. (2012). The Future of Accreditation. *Planning For Higher Education*, 40(3), 8-15.
- Elder, L., & Paul, R. (2003). Critical Thinking: Teaching Students How To Study and Learn (Part IV). *Journal Of Developmental Education*, 27(1), 36-37.
- Elder, L., & Paul, R. (2001). Critical Thinking: Thinking to Some Purpose. *Journal Of Developmental Education*, 25(1), 40-41.
- Ewell, P. T., & Jones, D. P. (2006). State-level accountability for higher education: On the edge of a transformation. *New Directions For Higher Education*, (135), 9-16. doi:10.1002/he.222
- Fielding, G. D., & Kameenui, E. & Gersten, R. (1983). A Comparison of an Inquiry and a Direct Instruction Approach to Teaching Legal Concepts and Applications to Secondary School Students. *Journal Of Educational Research*, 76(5), 287-93.
- Findlow, S. (2008). Accountability and innovation in higher education: a disabling tension?. *Studies In Higher Education*, 33(3), 313-329. doi:10.1080/03075070802049285
- Fitzpatrick, J., Christie, C., & Mark, M.M., (2009). *Evaluation in Action, Interviews with Expert Evaluators*. Thousand Oaks, CA: SAGE Publications
- Fitzpatrick, J., Sanders, J., & Worthen, B., (2011). *Program Evaluation: Alternative Approaches and Practical Guidelines* (4<sup>th</sup> Edition). New York: Longman.
- Fletcher, R., Meyer, L., Anderson, H., Johnston, P., & Rees, M. (2012). Faculty and Students Conceptions of Assessment in Higher Education. *Higher Education*, 64(1), 119-133. doi:10.1007/s10734-011-9484-1
- Flinders, D. J. (2005). The costs of accountability. *Journal Of Curriculum Studies*, 37(5), 621-629. doi:10.1080/00220270500038461
- Gall, M.D., Gall, J.P, & Borg, W.R. (2006). *Educational Research: An Introduction* (8<sup>th</sup> Edition). Upper Saddle River, NJ: Allyn & Bacon/Pearson.

- Ginsberg, R. & Levine, A. (2013). New Standards Will Demand a Lot More of Teacher Education. *The Chronicle of Higher Education*. Retrieved from <http://chronicle.com/article/Demanding-a-Lot-More-of/142275/> October 18<sup>th</sup>, 2013.
- Glover, D., & Levacic, R. (2007). *Educational Resource Management: An International Perspective*. London, England: Institute of Education Press.
- Hartle, T. W. (2012). Accreditation and the Public Interest: Can Accreditors Continue to Play a Central Role in Public Policy?. *Planning For Higher Education*, 40(3), 16-21.
- Hecker, D. Employment Outlook 2004-13: Occupational Employment Projections to 2014. *Monthly Labor Review*, 2005, 128(11), 70-101.
- Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, 6, 1-55.
- Hubbell, L. (2007). Quality, Efficiency, and Accountability: Definitions and Applications. *New Directions For Higher Education*, (140), 5-13.
- Imazeki, J., & Reschovsky, A. (2005). Assessing the Use of Econometric Analysis in Estimating the Costs of Meeting State Education Accountability Standards" Lessons From Texas. *Peabody Journal Of Education (0161956X)*, 80(3), 96-125. doi:10.1207/s15327930pje8003\_6
- Jamieson, V., & Azzam, T. (2012). The Use of Technology in Evaluation Practice. *Journal Of MultiDisciplinary Evaluation*, 8(18), 1-15. Retrieved from [http://journals.sfu.ca/jmde/index.php/jmde\\_1/article/view/340/345](http://journals.sfu.ca/jmde/index.php/jmde_1/article/view/340/345)
- Joint Committee on Standards for Educational Evaluation (JCSEE). (2011). Program Evaluation Standards. <http://www.jcsee.org/program-evaluation-standards/program-evaluation-standards-statements>
- Kallison Jr., J. M., & Cohen, P. (2010). A New Compact for Higher Education: Funding and Autonomy for Reform and Accountability. *Innovative Higher Education*, 35(1), 37-49. doi:10.1007/s10755-009-9123-2
- Karantzas, G. C., Avery, M. R., MacFarlane, S., Mussap, A., Tooley, G., Hazelwood, Z., & Fitness, J. (March 01, 2013). Enhancing critical analysis and problem-solving skills in undergraduate psychology: An evaluation of a collaborative learning and problem-based learning approach. *Australian Journal of Psychology*, 65, 1, 38-45.



- Kentucky Department of Education (KDE). [www.education.ky.gov](http://www.education.ky.gov).
- King, R. (2007). Governance and accountability in the higher education regulatory state. *Higher Education*, 53(4), 411-430. doi:10.1007/s10734-005-3128-2
- King, J., McKegg, K., Oakden, J., & Wehipeihana, N. (2013). Rubrics: A Method for Surfacing Values and Improving the Credibility of Evaluation. *Journal Of MultiDisciplinary Evaluation*, 9(21), 11-20. Retrieved from [http://journals.sfu.ca/jmde/index.php/jmde\\_1/article/view/374](http://journals.sfu.ca/jmde/index.php/jmde_1/article/view/374)
- Kiss, K. L., & Townsend, J. S. (2012). Teacher Inquiry from Knowledge to Knowledges. *Issues In Teacher Education*, 21(2), 23-41.
- Kirchner, A. (2012). Evaluation of Electronic Assessment Systems and Their Ability to Meet NCATE Standard 2. *Dissertations*. Paper 20. <http://digitalcommons.wku.edu/diss/20>
- Kline, R. B. (2011). *Principles and practice of structural equation modeling* (3<sup>rd</sup> ed.). New York: Guilford Press.
- Lynch, D. C., Greer, A. G., Larson, L. C., Cummings, D. M., Harriett, B. S., Dreyfus, K., & Clay, M. C. (2003). DESCRIPTIVE METAEVALUATION. *Evaluation & The Health Professions*, 26(4), 447-461.
- Madaus, G. F., & Stufflebeam, D. L. (1984). Educational Evaluation and Accountability: A Review of Quality Assurance Efforts. *American Behavioral Scientist*, 27(5), 649-72.
- MacCallum, R. C., Browne, M. W., & Sugawara, H. M. (1996). Power analysis and determination of sample size for covariance structure modeling. *Psychological Methods*, 1, 130-149.
- Marchand, S., & Stoner, J. (2012). A Brief History of Accountability in Higher Education. *Phi Kappa Phi Forum*, 92(1), 16-18.
- Marzano, R. J. (1998). What Are the General Skills of Thinking and Reasoning and How Do You Teach Them?. *Clearing House*, 71(5), 268-73.
- Mehta, J. (2013). How Paradigms Create Politics: The Transformation of American Educational Policy, 1980-2001. *American Educational Research Journal*, 50(2), 285-324.
- Mero, N. P., Guidice, R. M., & Anna, A. L. (2006). The Interacting Effects of Accountability and Individual Differences on Rater Response to a Performance-Rating Task. *Journal of Applied Psychology*, 36(4), 795-819.

- Mitchell, V. J. (1998). *A metaevaluation of the texas education agency's district effectiveness and compliance review*. (Order No. 9915183, Texas A&M University). *ProQuest Dissertations and Theses*, 233-233 p. Retrieved from <http://search.proquest.com/docview/304481192?accountid=14665>. (304481192).
- National Council for Accreditation of Teacher Education (NCATE) Standards (2008). <http://www.ncate.org/Standards/UnitStandards/tabid/123/Default.aspx>
- Ockerman, M. S., & Mason, E. M. (2012). Developing School Counseling Students' Social Justice Orientation through Service Learning. *Journal Of School Counseling*, 10(5)
- Pallas, A. (2011). Assessing the Future of Higher Education. *Society*, 48(3), 213-215. doi:10.1007/s12115-011-9419-6
- Paul-Elder critical thinking model* (2008). University of Louisville i2a website. Retrieved from <http://louisville.edu/ideastoaction/about/criticalthinking/framework>.
- Paul, R., & Elder, L. (2002). Critical Thinking: Teaching Students How To Study and Learn (Part I). *Journal Of Developmental Education*, 26(1), 36-37.
- Paul, R., & Elder, L. (2005). *A guide for educators to critical thinking competency standards*. Retrieved from <http://www.udlap.mx/promueve/ciedd/CR/pensamiento/criticalthinkingcompetencies.pdf>
- Payne, M., & Miller, M. (2009). A Collaborative Approach to Assessment: The Assessment and Improvement Management System (AIMS). *Issues In Teacher Education*, 18(1), 149-160.
- Professional Growth and Effectiveness System (2013). Kentucky Department of Education. <http://education.ky.gov/teachers/hieffteach/pages/designing-pges.aspx>
- Raudenbush, S. W., Bryk, A. S., Cheong, Y. F., & Congdon, R. T., Jr. (2000). *HLM6: Hierarchical linear and nonlinear modeling*. Lincolnwood, IL: Scientific Software.
- Rothwell, W.J., & Sullivan, R. (2005). *Practicing Organization Development* (2<sup>nd</sup> ed.). San Francisco, CA: John Wiley & Sons.
- Royal, K. (2010). Evaluating Faculty Perceptions of Student Learning Outcomes: A Rasch Measurement Analysis. *Journal Of MultiDisciplinary Evaluation*, 6(14), 18-31. Retrieved from [http://journals.sfu.ca/jmde/index.php/jmde\\_1/article/view/269](http://journals.sfu.ca/jmde/index.php/jmde_1/article/view/269)




- Sanders, J., & Nafziger, D. (2011). A Basis for Determining the Adequacy of Evaluation Designs. *Journal Of MultiDisciplinary Evaluation*, 7(15), 44- 78. Retrieved from [http://journals.sfu.ca/jmde/index.php/jmde\\_1/article/view/296/293](http://journals.sfu.ca/jmde/index.php/jmde_1/article/view/296/293)
- Schuh, J. H. (2003). Selected Accountability and Assessment Issues. *New Directions For Student Services*, (103), 87.
- Scriven, M. (1969). *An Introduction to Meta-Evaluation*. Educ Prod Rep.
- Scriven, M. (2010). Evaluation Bias and its Control\*. *Journal Of MultiDisciplinary Evaluation*, 7(15), 79-98. Retrieved from [http://journals.sfu.ca/jmde/index.php/jmde\\_1/article/view/299/294](http://journals.sfu.ca/jmde/index.php/jmde_1/article/view/299/294)
- Seraphin, K., Philippoff, J., Parisky, A., Degnan, K., & Warren, D. (2013). Teaching Energy Science as Inquiry: Reflections on Professional Development as a Tool to Build Inquiry Teaching Skills for Middle and High School Teachers. *Journal Of Science Education And Technology*, 22(3), 235-251.
- Shore, B. M., Chichekian, T., Syer, C. A., Aulls, M. W., & Frederiksen, C. H. (2012). Planning, Enactment, and Reflection in Inquiry-Based Learning: Validating the McGill Strategic Demands of Inquiry Questionnaire. *International Journal Of Science And Mathematics Education*, 10(2), 315-337.
- Shulman, L. S. (2000). Teacher development: Roles of domain expertise and pedagogical development. *Journal of Applied Developmental Psychology*, 21, 129-135.
- Shulman, L. S. (2002). Truth and Consequences? Inquiry and Policy in Research on Teacher Education. *Journal Of Teacher Education*, 53, 248-253.
- Shulman, L. S. (2006, April). *Pedagogies of uncertainty: Teaching for understanding, judgment and commitment*. Paper delivered on April 18, 2006, at the annual Grawemeyer Award in Education presentation, Louisville, KY. Digital streaming video retrieved on August 15, 2006, from <http://www.louisville.edu/television/grawemeyer.shulman.asx>.
- Shulman, L. S., & Shulman, J. H. (2004). How and What Teachers Learn: A Shifting Perspective. *Journal of Curriculum Studies*, 36, 257-271.
- Sibolski, E. H. (2012). What's an Accrediting Agency Supposed to Do? Institutional Quality and Improvement vs. Regulatory Compliance. *Planning For Higher Education*, 40(3), 22-28.
- Singer, F., & Voica, C. (2013). A Problem-Solving Conceptual Framework and Its Implications in Designing Problem-Posing Tasks. *Educational Studies In Mathematics*, 83(1), 9-26.

- Soundarajan, N. (2004). Program Assessment and Program Improvement: Closing the Loop. *Assessment & Evaluation In Higher Education*, 29(5), 597-610.
- Spanier, G. B. (2010). Creating Adaptable Universities. *Innovative Higher Education*, 35(2), 91-99. doi:10.1007/s10755-009-9134-z
- Stake, R.E. (1995). *The art of case study research*. Thousand Oaks, CA: Sage.
- Stevens, J. (2009). *Applied multivariate statistics for the social sciences* (5th ed.). New York: Routledge.
- Stufflebeam, D. L. (2001). Evaluation Checklists: Practical Tools for Guiding and Judging Evaluations. *American Journal Of Evaluation*, 22(1), 71-79.
- Stufflebeam, D. L. (2001). The Metaevaluation Imperative. *American Journal of Evaluation*, 22(2), 183-209.
- Stufflebeam, D. L. (1999). Program Evaluation Models Metaevaluation Checklist. Retrieved from [www.wmich.edu/evalctr/checklists](http://www.wmich.edu/evalctr/checklists)
- Stufflebeam, D. L. (2011). Meta-Evaluation. *Journal Of Multidisciplinary Evaluation*, 7(15), 99-158.
- Taylor-Ritzler, T., Suarez-Balcazar, Y., Garcia-Iriarte, E., Henry, D. B., & Balcazar, F. E. (2013). Understanding and Measuring Evaluation Capacity: A Model and Instrument Validation Study. *American Journal Of Evaluation*, 34(2), 190-206. doi:10.1177/1098214012471421
- Teaching, Empowering, Leading and Learning (TELL) Kentucky (2013). Kentucky Department of Education (KDE). [www.tellkentucky.org](http://www.tellkentucky.org)
- Thelin, J. R. (2011). *A History of American Higher Education* (2<sup>nd</sup> ed.) . Baltimore, MD: The Johns Hopkins University Press.
- Toma, J. (2008). A few reflections on accountability, affordability, and access worldwide. *New Directions For Institutional Research*, 200893-97.
- University of Louisville I2A Website. Retrieved from <http://louisville.edu/ideastoaction/i2a-uofl/critical-thinking/paulelder-critical-thinking-model/paul-elder-model-of-critical-thinking.html>.
- University of Louisville 2013-2014 Just the Facts, Retrieved from <https://louisville.edu/institutionalresearch/key-institutional-data>

- U.S. Department of Education. <http://ope.ed.gov/accreditation/>. Retrieved March, 28<sup>th</sup>, 2013.
- U.S. Department of Education (2001). *No Child Left Behind*.  
<http://www2.ed.gov/policy/elsec/leg/esea02/index.html>
- U.S. Department of Education (2010). *College- and Career-Ready Students*.  
[http://www2.ed.gov/policy/elsec/leg/blueprint/publication\\_pg4.html](http://www2.ed.gov/policy/elsec/leg/blueprint/publication_pg4.html)
- Vanhoof, J., & Van Petegem, P. (2010). Evaluating the Quality of Self-Evaluations: The (Mis)Match between Internal and External Meta-Evaluation. *Studies In Educational Evaluation*, 36(1-2), 20-26.
- Webber, K. L., & Boehmer, R. G. (2008). The balancing act: Accountability, affordability, and access in American higher education. *New Directions For Institutional Research*, 200879-91.
- Wentling, T. L., & Klit, J. A. (1973). *Meta-Evaluation Applied: The Evaluation of a Large Scale Evaluation System*.
- Wilkins, E. A., Young, A., & Sterner, S. (2009). An Examination of Institutional Reports: Use of Data for Program Improvement. *Action In Teacher Education*, 31(1), 14-23.
- Williams, J. M., & Greenleaf, A. T. (2012). Ecological Psychology: Potential Contributions to Social Justice and Advocacy in School Settings. *Journal Of Educational & Psychological Consultation*, 22(1-2), 141-157.
- Wingate, L. A. (2009). *The program evaluation standards applied for metaevaluation purposes: Investigating interrater reliability and implications for use*. (Order No. 3392166, Western Michigan University). *ProQuest Dissertations and Theses*, , 181. Retrieved from <http://search.proquest.com/docview/305029057?accountid=14665>. (305029057).
- Zemsky, R., Wegner, G. R., & Massy, W. F., (2005). *Remaking the American university: Market-smart and mission-centered*. Piscataway, NJ: Rutgers University Press.

APPENDIX A

Sample Continuous Assessment Record and Documentation System (CARDS) Chart.

|                        |   |   |   |
|------------------------|--|---|--|
| Criteria               | CARDS 1<br>Admission   | CARDS 2<br>Pre-clinical /Mid-<br>point  | CARDS 3<br>Clinical<br>Practice/<br>Completion   |
| Required Check-Points: | <p><b>Admissions Check-Points:</b></p> <p><b>Written Communication:</b><br/>ENG 102 or equivalent (C or above)</p> <p><b>Oral Communication:</b><br/>Speech communication COM 115 or equivalent (C or above) or speech proficiency exam;</p> <p><b>3 letters of Recommendations</b><br/>(Academic/Faculty, Professional, and Work with Children)</p> <p><b>Statement of Understanding of Admissions Guidelines</b><br/>Signed statement in application</p> | <p><b>Mid-Program Check-Points:</b></p> <p><b>Field Experience Required Checks:</b><br/>Background Check /TB</p> <p><b>Student Teaching Required Checks:</b><br/>State Criminal Check/TB<br/>Medical/Federal Criminal Check /Insurance</p> <p>Satisfactory mid-point portfolio</p> <p>Positive recommendation from the Elementary Program Committee</p> | <p><b>Program Completion Check-Points:</b></p> <p><b>Degree Check</b></p> <p><b>Graduation Application (EASS) TC 1 Completed</b></p> |

|  |   |   |  |
|--|---|---|--|
|  | <p><b>Academic Program Sheet</b> (must be signed by advisor and candidate and submitted to the Education Advising Student Services)</p> <p><b>Character and Fitness Form</b></p>  |   |  |
| <p><b>21<sup>st</sup> Century Skills (Critical Thinking, Collaboration, Communication, and Creativity)</b></p> | <p><b>Critical Thinking</b><br/>Ideas to Action<br/>Holistic Construct<br/>Rubric<br/>(Professional Statement, Interview, and Letters of Recommendation)</p> <p><b>Assessment of Collaboration</b><br/>(Based on Interview Question 5 and/or Professional Statement)</p> <p><b>UofL Effective Communication Rubric</b> (Professional Statement, Letters of Recommendation, and Interview)</p> <p><b>Assessment of Creativity</b><br/>(Professional Statement and Interview)</p> |   |  |
| <p><b>Orientations</b></p>   | <p>Candidates are required to attend a Program Orientation upon admission</p>   | <p><b>Student Teaching Orientations</b></p> |  |

|  |   |  |   |
|--|---|--|---|
| <p><b>Academic Content and Professional Knowledge GPA and Minimum Credit Hours</b></p> | <p><b>GPA:</b> Suggested minimum cumulative 2.75 OR A grade point average of 3.00 on a 4.0 scale on the last thirty (30) hours of credit completed; and</p> <p>Cumulative Pre-professional GPA is 3.0 or higher for the following courses: EDTP 201, EDTP 107, MATH 151 and MATH 152</p> <p>45 Semester Credit Hours (UG)</p> | <p><b>GPA:</b> Cumulative 2.75 Professional 3.0 (Suggested Minimums)</p> <p>Completion of required courses on program sheet, with required GPA. See program sheet for specifics.</p> | <p><b>GPA:</b> Cumulative 2.75 Professional 3.0 (Suggested Minimums)</p>  |
| <p><b>Academic Competency – Content Knowledge</b></p>                                  | <p><b>Academic Competency:</b> Suggested minimum PPST scores (R-176, M-174, W-174)</p>  |  | <p><b>Praxis II:</b> Elementary Praxis Content Exams</p> <p>PLT Exam</p>  |
| <p><b>Conceptual Framework Constructs</b></p>  | <p><b>Ideas to Action Holistic Construct Rubric –See also under 21<sup>st</sup> Century Skills)</b> (Professional Statement, Letters of Recommendation, Interview, etc.)</p>  | <p><b>Ideas to Action Holistic Construct Rubric</b> (Hallmark Assessments and Rationale provided in the mid-program portfolio)</p>   | <p><b>Ideas to Action Holistic Construct Rubric</b> (Hallmark Assessments and Rationale provided in the exit portfolio)</p> |
| <p><b>Field and Clinical Placements</b></p>  |   | <p><b>Field Hours-</b> Minimum of 200 hours (UG)</p> <p>Field Hours documented in EPSB KFETS System</p>  | <p><b>Student Teaching Observation Forms</b> 4 formal observations by the university supervisor.</p>                        |



|  |   |  |   |
|--|---|--|---|
|  |   |  |   |
| <b>Conceptual Framework Dispositions</b> | <b>Ideas to Action Unit Dispositions Rubric</b><br>(Professional Statement, Interview, Letters of Recommendation) | <b>Ideas to Action Unit Dispositions Rubric</b><br><br>Mid-Program Portfolio<br><br>Satisfactory Dispositions Assessment for Candidates Completing Content and Special Methods from Mentor Teachers, Supervisor and/or Instructors | <b>Ideas to Action Unit Dispositions Rubric</b><br><br>Exit Portfolio<br><br>Student Teacher Candidate Dispositions Assessment from Cooperating Teachers, Supervisor and/or Instructors |
| <b>Code of Ethics</b>                    | <b>Kentucky Code of Ethics</b><br>Signed statement in application   | <b>Kentucky Code of Ethics</b><br>Student Teaching Orientation   |   |
| <b>Technology</b>                        | Signed Acceptable Use of Technology Agreement   | <b>Unit Assessment for Technology</b><br>(Kentucky Teacher Standard 6 assessed in Mid-Program Portfolio)   | <b>Unit Assessment for Technology</b><br>(Kentucky Teacher Standard 6 assessed in Exit Portfolio)   |
| <b>Diversity</b>                         | Interview Question 3  | <b>Unit Assessment for Diversity</b><br>(UofL Standard 11 assessed in Mid-Program Portfolio)   | <b>Unit Assessment for Diversity</b><br>(UofL Standard 11 assessed in Exit Portfolio)   |
| <b>Evidence of Planning</b>              |   | <b>Unit Assessment for Evidence of Planning</b><br>(Kentucky Teacher Standard 2 assessed in Mid-Program Portfolio)   | <b>Unit Assessment for Evidence of Planning</b><br>(Kentucky Teacher Standard 2 assessed in Exit Portfolio)   |
| <b>Impact on P-12 Student Learning</b>   |   |  | <b>Unit Assessment for Impact on P-12 Student</b>   |

|                  |  |  |  |
|------------------|--|--|--|
|                  |  |  | <b>Learning</b><br>(Student Teaching Instructional Unit – EDTP 477)  |
| <b>Portfolio</b> | <b>CARDS 1 Interview:</b><br>Program Faculty and School Partners | <b>CARDS 2 Portfolio:</b><br>Kentucky Teacher Standards (KTS) (½ standards,) and Letter to Reader, Statement of Authenticity | <b>CARDS 3 Portfolio:</b><br>Kentucky Teacher Standards (KTS) (all standards), revised Letter to Reader, Statement of Authenticity |

## APPENDIX B

### JCSEE Program Evaluation Standards (2011)

---

| <b>Standards</b>      |   |
|-----------------------|---|
| Utility Standards     | <p>The utility standards are intended to increase the extent to which program stakeholders find evaluation processes and products valuable in meeting their needs.</p> <p><b>U1 Evaluator Credibility</b> Evaluations should be conducted by qualified people who establish and maintain credibility in the evaluation context.</p> <p><b>U2 Attention to Stakeholders</b> Evaluations should devote attention to the full range of individuals and groups invested in the program and affected by its evaluation.</p> <p><b>U3 Negotiated Purposes</b> Evaluation purposes should be identified and continually negotiated based on the needs of stakeholders.</p> <p><b>U4 Explicit Values</b> Evaluations should clarify and specify the individual and cultural values underpinning purposes, processes, and judgments.</p> <p><b>U5 Relevant Information</b> Evaluation information should serve the identified and emergent needs of stakeholders.</p> <p><b>U6 Meaningful Processes and Products</b> Evaluations should construct activities, descriptions, and judgments in ways that encourage participants to rediscover, reinterpret, or revise their understandings and behaviors.</p> <p><b>U7 Timely and Appropriate Communicating and Reporting</b> Evaluations should attend to the continuing information needs of their multiple audiences.</p> <p><b>U8 Concern for Consequences and Influence</b> Evaluations should promote responsible and adaptive use while guarding against unintended negative consequences and misuse.</p> |
| Feasibility Standards | <p>The feasibility standards are intended to increase evaluation effectiveness and efficiency.</p> <p><b>F1 Project Management</b> Evaluations should use effective project management strategies.</p> <p><b>F2 Practical Procedures</b> Evaluation procedures should be practical and responsive to the way the program operates.</p> <p><b>F3 Contextual Viability</b> Evaluations should recognize, monitor, and balance the cultural and political interests and needs of individuals and groups.</p> <p><b>F4 Resource Use</b> Evaluations should use resources effectively and</p>  |

|                     |  |
|---------------------|--|
| Propriety Standards | <p>efficiently.</p> <p>The propriety standards support what is proper, fair, legal, right and just in evaluations.</p> <p><b>P1 Responsive and Inclusive Orientation</b> Evaluations should be responsive to stakeholders and their communities.</p> <p><b>P2 Formal Agreements</b> Evaluation agreements should be negotiated to make obligations explicit and take into account the needs, expectations, and cultural contexts of clients and other stakeholders.</p> <p><b>P3 Human Rights and Respect</b> Evaluations should be designed and conducted to protect human and legal rights and maintain the dignity of participants and other stakeholders.</p> <p><b>P4 Clarity and Fairness</b> Evaluations should be understandable and fair in addressing stakeholder needs and purposes.</p> <p><b>P5 Transparency and Disclosure</b> Evaluations should provide complete descriptions of findings, limitations, and conclusions to all stakeholders, unless doing so would violate legal and propriety obligations.</p> <p><b>P6 Conflicts of Interests</b> Evaluations should openly and honestly identify and address real or perceived conflicts of interests that may compromise the evaluation.</p> <p><b>P7 Fiscal Responsibility</b> Evaluations should account for all expended resources and comply with sound fiscal procedures and processes.</p> |
| Accuracy Standards  | <p>The accuracy standards are intended to increase the dependability and truthfulness of evaluation representations, propositions, and findings, especially those that support interpretations and judgments about quality.</p> <p><b>A1 Justified Conclusions and Decisions</b> Evaluation conclusions and decisions should be explicitly justified in the cultures and contexts where they have consequences.</p> <p><b>A2 Valid Information</b> Evaluation information should serve the intended purposes and support valid interpretations.</p> <p><b>A3 Reliable Information</b> Evaluation procedures should yield sufficiently dependable and consistent information for the intended uses.</p> <p><b>A4 Explicit Program and Context Descriptions</b> Evaluations should document programs and their contexts with appropriate detail and scope for the evaluation purposes.</p> <p><b>A5 Information Management</b> Evaluations should employ systematic information collection, review, verification, and storage methods.</p> <p><b>A6 Sound Designs and Analyses</b> Evaluations should employ technically adequate designs and analyses that are appropriate for the evaluation purposes.</p> <p><b>A7 Explicit Evaluation Reasoning</b> Evaluation reasoning leading from information and analyses to findings, interpretations,</p>                   |

Evaluation  
Accountability  
Standards

conclusions, and judgments should be clearly and completely documented.

**A8 Communication and Reporting** Evaluation communications should have adequate scope and guard against misconceptions, biases, distortions, and errors.

The evaluation accountability standards encourage adequate documentation of evaluations and a metaevaluative perspective focused on improvement and accountability for evaluation processes and products.

**E1 Evaluation Documentation** Evaluations should fully document their negotiated purposes and implemented designs, procedures, data, and outcomes.

**E2 Internal Metaevaluation** Evaluators should use these and other applicable standards to examine the accountability of the evaluation design, procedures employed, information collected, and outcomes.

**E3 External Metaevaluation** Program evaluation sponsors, clients, evaluators, and other stakeholders should encourage the conduct of external metaevaluations using these and other applicable standards.

---

APPENDIX C

One-Way Random Effects ANOVA Results

Table B.1

*One-Way Random Effects ANOVA for Inquiry*

| Fixed effects                                 | Coefficient (SE) | <i>t</i> (df) | <i>p</i> |          |
|---|------------------|---------------|----------|----------|
| Model for intercept Inquiry (b <sub>0</sub> ) |                  |               |          |          |
| Intercept (g <sub>00</sub> )                  | 2.73 (.04)       | 74.62 (37)    | <.001    |          |
| Random effects                                |                  |               |          |          |
|   | Variance         | Df            | $\chi^2$ | <i>p</i> |
| Var. in intercepts (t <sub>00</sub> )         | .029             | 37            | 125.22   | <.001    |
| Var. within programs (s <sup>2</sup> )        | .17              |               |          |          |

Table B.2

*One-Way Random Effects ANOVA for Action*

| Fixed effects                                | Coefficient (SE) | <i>t</i> (df) | <i>p</i> |          |
|--|------------------|---------------|----------|----------|
| Model for intercept Action (b <sub>0</sub> ) |                  |               |          |          |
| Intercept (g <sub>00</sub> )                 | 2.67 (.04)       | 76.18 (37)    | <.001    |          |
| Random effects                               |                  |               |          |          |
|  | Variance         | Df            | $\chi^2$ | <i>p</i> |
| Var. in intercepts (t <sub>00</sub> )        | .024             | 37            | 101.58   | <.001    |
| Var. within programs (s <sup>2</sup> )       | .20              |               |          |          |

Table B.3

*One-Way Random Effects ANOVA for Advocacy*

| Fixed effects                   | Coefficient (SE) | <i>t</i> (df) | <i>p</i> |          |
|---------------------------------|------------------|---------------|----------|----------|
| Model for intercept             |                  |               |          |          |
| Advocacy ( $b_0$ )              |                  |               |          |          |
| Intercept ( $g_{00}$ )          | 2.64 (.04)       | 69.595 (37)   | <.001    |          |
| Random effects                  | Variance         | Df            | $\chi^2$ | <i>p</i> |
| Var. in intercepts ( $t_{00}$ ) | .029             | 37            | 112.80   | <.001    |
| Var. within programs ( $s^2$ )  | .20              |               |          |          |

Table B.4

*One-Way Random Effects ANOVA for Inquiry Disposition*

| Fixed effects                                     | Coefficient (SE) | <i>t</i> (df) | <i>P</i> |          |
|---|------------------|---------------|----------|----------|
| Model for intercept Inquiry Disposition ( $b_0$ ) |                  |               |          |          |
| Intercept ( $g_{00}$ )                            | 2.73 (.04)       | 747.61 (37)   | <.001    |          |
| Random effects                                    | Variance         | Df            | $\chi^2$ | <i>p</i> |
| Var. in intercepts ( $t_{00}$ )                   | .026             | 37            | 110.40   | <.001    |
| Var. within programs ( $s^2$ )                    | .18              |               |          |          |

Table B.5

*One-Way Random Effects ANOVA for Action Disposition*

| Fixed effects                                    | Coefficient (SE) | <i>t</i> (df) | <i>p</i> |          |
|--|------------------|---------------|----------|----------|
| Model for intercept Action Disposition ( $b_0$ ) |                  |               |          |          |
| Intercept ( $g_{00}$ )                           | 2.73 (.04)       | 82.812 (37)   | <.001    |          |
| Random effects                                   | Variance         | Df            | $\chi^2$ | <i>p</i> |
| Var. in intercepts ( $t_{00}$ )                  | .02              | 37            | 93.55    | <.001    |
| Var. within programs ( $s^2$ )                   | .18              |               |          |          |

Table B.6

*One-Way Random Effects ANOVA for Advocacy Disposition*

| Fixed effects                                      | Coefficient (SE) | <i>t</i> (df) | <i>p</i> |          |
|--|------------------|---------------|----------|----------|
| Model for intercept Advocacy Disposition ( $b_0$ ) |                  |               |          |          |
| Intercept ( $g_{00}$ )                             | 2.68 (.03)       | 77.59 (37)    | <.001    |          |
| Random effects                                     | Variance         | Df            | $\chi^2$ | <i>p</i> |
| Var. in intercepts ( $t_{00}$ )                    | .022             | 37            | 94.70    | <.001    |
| Var. within programs ( $s^2$ )                     | .20              |               |          |          |



Table B.7

*One-Way Random Effects ANOVA for Impact*

| Fixed effects                                | Coefficient (SE) | <i>t</i> (df) | <i>p</i> |          |
|--|------------------|---------------|----------|----------|
| Model for intercept Impact (b <sub>0</sub> ) |                  |               |          |          |
| Intercept (g <sub>00</sub> )                 | 2.79 (.02)       | 116.39 (37)   | <.001    |          |
| Random effects                               | Variance         | Df            | $\chi^2$ | <i>p</i> |
| Var. in intercepts (t <sub>00</sub> )        | .005             | 37            | 49.98    | .075     |
| Var. within programs (s <sup>2</sup> )       | .19              |               |          |          |

Table B.8

*One-Way Random Effects ANOVA for Cycle1*

| Fixed effects                                | Coefficient (SE) | <i>t</i> (df) | <i>p</i> |          |
|--|------------------|---------------|----------|----------|
| Model for intercept Cycle1 (b <sub>0</sub> ) |                  |               |          |          |
| Intercept (g <sub>00</sub> )                 | 2.70 (.04)       | 60.89 (37)    | <.001    |          |
| Random effects                               | Variance         | Df            | $\chi^2$ | <i>p</i> |
| Var. in intercepts (t <sub>00</sub> )        | .051             | 37            | 191.46   | <.001    |
| Var. within programs (s <sup>2</sup> )       | .17              |               |          |          |

Table B.9

*One-Way Random Effects ANOVA for Cycle2*

| Fixed effects                                | Coefficient (SE) | <i>t</i> (df) | <i>p</i> |          |
|--|------------------|---------------|----------|----------|
| Model for intercept Cycle2 (b <sub>0</sub> ) |                  |               |          |          |
| Intercept (g <sub>00</sub> )                 | 2.78 (.04)       | 79.40 (37)    | <.001    |          |
| Random effects                               | Variance         | Df            | $\chi^2$ | <i>p</i> |
| Var. in intercepts (t <sub>00</sub> )        | .028             | 37            | 130.03   | <.001    |
| Var. within programs (s <sup>2</sup> )       | .15              |               |          |          |

Table B.10

*One-Way Random Effects ANOVA for Cycle3*

| Fixed effects                                | Coefficient (SE) | <i>t</i> (df) | <i>p</i> |          |
|--|------------------|---------------|----------|----------|
| Model for intercept Cycle3 (b <sub>0</sub> ) |                  |               |          |          |
| Intercept (g <sub>00</sub> )                 | 2.85 (.02)       | 129.75 (37)   | <.001    |          |
| Random effects                               | Variance         | Df            | $\chi^2$ | <i>p</i> |
| Var. in intercepts (t <sub>00</sub> )        | .006             | 37            | 57.36    | .017     |
| Var. within programs (s <sup>2</sup> )       | .12              |               |          |          |

Table B.11

*One-Way Random Effects ANOVA for Cycle4*

| Fixed effects                                | Coefficient (SE) | <i>t</i> (df) | <i>p</i> |          |
|--|------------------|---------------|----------|----------|
| Model for intercept Cycle4 (b <sub>0</sub> ) |                  |               |          |          |
| Intercept (g <sub>00</sub> )                 | 2.93 (.01)       | 205.78 (37)   | <.001    |          |
| Random effects                               | Variance         | Df            | $\chi^2$ | <i>p</i> |
| Var. in intercepts (t <sub>00</sub> )        | .002             | 37            | 46.30    | .141     |
| Var. within programs (s <sup>2</sup> )       | .06              |               |          |          |

Table B.12

*One-Way Random Effects ANOVA for KTS1*

| Fixed effects                              | Coefficient (SE) | <i>t</i> (df) | <i>p</i> |          |
|--|------------------|---------------|----------|----------|
| Model for intercept KTS1 (b <sub>0</sub> ) |                  |               |          |          |
| Intercept (g <sub>00</sub> )               | 2.80 (.03)       | 87.28 (37)    | <.001    |          |
| Random effects                             | Variance         | Df            | $\chi^2$ | <i>p</i> |
| Var. in intercepts (t <sub>00</sub> )      | .022             | 37            | 117.04   | <.001    |
| Var. within programs (s <sup>2</sup> )     | .13              |               |          |          |

Table B.13

*One-Way Random Effects ANOVA for KTS2*

| Fixed effects                              | Coefficient (SE) | <i>t</i> (df) | <i>p</i> |          |
|--|------------------|---------------|----------|----------|
| Model for intercept KTS2 (b <sub>0</sub> ) |                  |               |          |          |
| Intercept (g <sub>00</sub> )               | 2.74 (.03)       | 82.53 (37)    | <.001    |          |
| Random effects                             | Variance         | Df            | $\chi^2$ | <i>p</i> |
| Var. in intercepts (t <sub>00</sub> )      | .021             | 37            | 97.11    | <.001    |
| Var. within programs (s <sup>2</sup> )     | .17              |               |          |          |

Table B.14

*One-Way Random Effects ANOVA for KTS3*

| Fixed effects                              | Coefficient (SE) | <i>t</i> (df) | <i>p</i> |          |
|--|------------------|---------------|----------|----------|
| Model for intercept KTS3 (b <sub>0</sub> ) |                  |               |          |          |
| Intercept (g <sub>00</sub> )               | 2.79 (.03)       | 93.55 (37)    | <.001    |          |
| Random effects                             | Variance         | Df            | $\chi^2$ | <i>p</i> |
| Var. in intercepts (t <sub>00</sub> )      | .018             | 37            | 97.58    | <.001    |
| Var. within programs (s <sup>2</sup> )     | .14              |               |          |          |

Table B.15

*One-Way Random Effects ANOVA for KTS4*

| Fixed effects                              | Coefficient (SE) | <i>t</i> (df) | <i>p</i> |          |
|--|------------------|---------------|----------|----------|
| Model for intercept KTS4 (b <sub>0</sub> ) |                  |               |          |          |
| Intercept (g <sub>00</sub> )               | 2.78 (.03)       | 92.89 (37)    | <.001    |          |
| Random effects                             | Variance         | Df            | $\chi^2$ | <i>p</i> |
| Var. in intercepts (t <sub>00</sub> )      | .017             | 37            | 100.03   | <.001    |
| Var. within programs (s <sup>2</sup> )     | .14              |               |          |          |

Table B.16

*One-Way Random Effects ANOVA for KTS5*

| Fixed effects                              | Coefficient (SE) | <i>t</i> (df) | <i>p</i> |          |
|--|------------------|---------------|----------|----------|
| Model for intercept KTS5 (b <sub>0</sub> ) |                  |               |          |          |
| Intercept (g <sub>00</sub> )               | 2.63 (.04)       | 63.28 (37)    | <.001    |          |
| Random effects                             | Variance         | Df            | $\chi^2$ | <i>p</i> |
| Var. in intercepts (t <sub>00</sub> )      | .040             | 37            | 134.68   | <.001    |
| Var. within programs (s <sup>2</sup> )     | .20              |               |          |          |

Table B.17

*One-Way Random Effects ANOVA for KTS6*

| Fixed effects                              | Coefficient (SE) | <i>t</i> (df) | <i>p</i> |          |
|--|------------------|---------------|----------|----------|
| Model for intercept KTS6 (b <sub>0</sub> ) |                  |               |          |          |
| Intercept (g <sub>00</sub> )               | 2.65 (.04)       | 65.25 (37)    | <.001    |          |
| Random effects                             | Variance         | Df            | $\chi^2$ | <i>p</i> |
| Var. in intercepts (t <sub>00</sub> )      | .039             | 37            | 138.09   | <.001    |
| Var. within programs (s <sup>2</sup> )     | .18              |               |          |          |

Table B.18

*One-Way Random Effects ANOVA for KTS7*

| Fixed effects                              | Coefficient (SE) | <i>t</i> (df) | <i>p</i> |          |
|--|------------------|---------------|----------|----------|
| Model for intercept KTS7 (b <sub>0</sub> ) |                  |               |          |          |
| Intercept (g <sub>00</sub> )               | 2.73 (.03)       | 82.04 (37)    | <.001    |          |
| Random effects                             | Variance         | Df            | $\chi^2$ | <i>p</i> |
| Var. in intercepts (t <sub>00</sub> )      | .021             | 37            | 97.07    | <.001    |
| Var. within programs (s <sup>2</sup> )     | .18              |               |          |          |

Table B.19

*One-Way Random Effects ANOVA for KTS8*

| Fixed effects                              | Coefficient (SE) | <i>t</i> (df) | <i>p</i> |          |
|--|------------------|---------------|----------|----------|
| Model for intercept KTS8 (b <sub>0</sub> ) |                  |               |          |          |
| Intercept (g <sub>00</sub> )               | 2.74 (.03)       | 86.03 (37)    | <.001    |          |
| Random effects                             | Variance         | Df            | $\chi^2$ | <i>p</i> |
| Var. in intercepts (t <sub>00</sub> )      | .019             | 37            | 90.90    | <.001    |
| Var. within programs (s <sup>2</sup> )     | .17              |               |          |          |

Table B.20

*One-Way Random Effects ANOVA for KTS9*

| Fixed effects                              | Coefficient (SE) | <i>t</i> (df) | <i>p</i> |          |
|--|------------------|---------------|----------|----------|
| Model for intercept KTS9 (b <sub>0</sub> ) |                  |               |          |          |
| Intercept (g <sub>00</sub> )               | 2.78 (.03)       | 82.22 (37)    | <.001    |          |
| Random effects                             | Variance         | Df            | $\chi^2$ | <i>p</i> |
| Var. in intercepts (t <sub>00</sub> )      | .025             | 37            | 119.54   | <.001    |
| Var. within programs (s <sup>2</sup> )     | .15              |               |          |          |

Table B.21

*One-Way Random Effects ANOVA for KTS10*

| Fixed effects                               | Coefficient (SE) | <i>t</i> (df) | <i>p</i> |          |
|---|------------------|---------------|----------|----------|
| Model for intercept KTS10 (b <sub>0</sub> ) |                  |               |          |          |
| Intercept (g <sub>00</sub> )                | 2.72 (.03)       | 75.20 (37)    | <.001    |          |
| Random effects                              | Variance         | Df            | $\chi^2$ | <i>p</i> |
| Var. in intercepts (t <sub>00</sub> )       | .028             | 37            | 116.88   | <.001    |
| Var. within programs (s <sup>2</sup> )      | .18              |               |          |          |

Table B.22

*One-Way Random Effects ANOVA for UofL11*

| Fixed effects                                | Coefficient (SE) | <i>t</i> (df) | <i>p</i> |          |
|--|------------------|---------------|----------|----------|
| Model for intercept UofL11 (b <sub>0</sub> ) |                  |               |          |          |
| Intercept (g <sub>00</sub> )                 | 2.69 (.04)       | 73.99 (37)    | <.001    |          |
| Random effects                               | Variance         | Df            | $\chi^2$ | <i>p</i> |
| Var. in intercepts (t <sub>00</sub> )        | .028             | 37            | 119.24   | <.001    |
| Var. within programs (s <sup>2</sup> )       | .18              |               |          |          |



## APPENDIX D

### Institutional Review Board

#### **Inquiry and Evidence: Self Study across the Division of Teaching and Learning, College of Education and Human Development (Working Title)**

##### **Contacts:**

**Dr. Shelley Thomas**

**Dr. Melissa Shirley**

**Dr. Nicole Fenty**

**Dr. Caroline Sheffield**

**Dr. Penny Howell**

**Dr. Christine Sherretz**

**Ms. Stefanie Wooten Burnett**

Our mission is to advance knowledge and understanding across our disciplines and constituencies and to develop educational leaders who will inform policy, improve practice, strengthen communities, and address pressing social concerns. We prepare students to be exemplary professional practitioners and scholars; to generate, use, and disseminate knowledge about teaching...

CEHD Mission Statement (August 2007).

##### **Summary**

The primary objective of the proposed study is to conduct ongoing self-assessment of the impact of course and other program experiences in the Division of Teaching and Learning within the College of Education and Human Development (CEHD). As teachers and scholars, self-study is an integral and embedded part of faculty work in the Division and a means to address the tenants in our Mission Statement.

Within an ongoing purpose to collect and analyze course impact, faculty who serve as both instructors and administrators in the Division may access both current and longitudinal data. The creation of a continuing, systematic self-study in the Division that is also IRB reviewed aids the fulfillment of important internal *and* external responsibilities of the institution.

The study will include prospective and retrospective evidence from division of teaching and learning courses. Specifically, the courses and program experiences are those offered by the departments of (a) Early Childhood and Elementary Education; (b) Middle and Secondary Education; (c) Special Education; and (d) Physical Education.

## Introduction

Historically, research on teacher preparation has addressed a fundamental question: “How do you make a teacher?” (Schwartz, 1996, p. 3). This question as well as evaluation of the relative effectiveness of particular frameworks and strategies for preparing teachers has been at the core of research on teaching as well as research on teacher education. Indeed, the roots of the preparation query originate with Plato and Socrates (Schwartz, 1996).

As alluded to, research on teaching and research on teacher preparation overlap a great deal. However, while similar, *research on teaching* and *research on teacher preparation or education* are distinct. For the purpose of this proposed study, we distinguish the two as do teacher educators, Cochran-Smith & Fries (2006). In their body of work, *research on teaching* refers to that work located in P-12 schools referencing learners who are P-12 pupils. *Research on teacher preparation or education* refers to work preparing P-12 teachers, referencing learners who include both preservice teachers<sup>1</sup> as well as those certified teachers returning to advance their educations. Thus, as teacher educators who work primarily in a College of Education, this self- study is situated within the former while drawing from and contributing to the latter.

### ***Studying Teacher Education: Context for the Work in the CEHD***

In the midst of debates about the effectiveness of different programs and routes for the preparation of teachers, the American Educational Research Association<sup>2</sup> initiated a group comprised of nationally recognized scholars in the field. This group, named the AERA Panel on Research and Teacher Education, formed in 1999. The panel’s work resulted in the 2006 publication *Studying Teacher Education: A Report of the AERA Panel on Research and Teacher Education*. The *Report* is contextualized within several issues and emerging questions with respect to teacher education. These originate with policy makers as well as the public at large. Questions and issues relevant to this self-study are referenced.

First, the merits of teacher education as a field historically reflect debates around the best ways to recruit, prepare, and retain teachers. Next, constituents across policy and practice contexts question the relationships between research on teacher preparation and both practice and policy within teacher education. Finally, one particularly crucial, challenging question: how do the teacher preparation experiences of candidates and practicing teachers influence pupil learning in the classroom (Cochran-Smith & Zeichner, 2005)?

---

<sup>1</sup> In the literature on teacher education and teacher accreditation, preservice teachers are also referred to as candidates.

<sup>2</sup> AERA is the most prominent international professional organization, with the primary goal of advancing educational research and its practical application

At Boston College, teacher educators participate in a multi-disciplinary, large-scale, longitudinal initiative titled Teachers for a New Era (TNE) (See <http://www.teachersforanewera.org/> ). The TNE work, in particular, serves as a model for the improvement of teacher education programming and as well as a site for empirical research in order to enrich and broaden the current teacher education knowledge base and improve teacher education. BC's portfolio of studies from the TNE work inform BC curricular decisions that support their own students as well as address the larger questions across teacher education as a field (Cochran-Smith, et al., 2009).

The self-study planned in Teaching and Learning will likewise address questions that effect practice within and across departments and programs in our institution and contribute to questions and issues in the field. Further, this work frames much of the current scholarship carried out by instructors in the Division of Teaching and Learning. By using the extensive data available from the assessment system, as well as other course artifacts within an IRB approved research study, faculty can employ empirical methods to investigate questions relevant to teacher education in general and in their own fields and specialties.

***Addressing Accreditation Standards: Commitment to Quality and Accountability***

This study will enable the Division of Teaching and Learning to address several standards necessary for accreditation purposes. Further, reports from the study provide “evidence” which may be submitted for accreditation reviews.

Relevant Accrediting Bodies and Standards, as well as other agencies requiring evidence of performance:

National Council for Accreditation of Teacher Education

Standard 1: *Candidate Knowledge, Skills, and Professional Dispositions*

Standard 2: *Assessment System and Unit Evaluation*

Standard 4: *Diversity*

Title 2 of the Higher Education Act. (see <https://title2.ed.gov/View.asp>)

Educational Professional Standards Board (EPSB; see

<http://www.kyepsb.net/certification/index.asp>)

**Study Objectives**

- 1) Enable faculty to systematically and routinely use evidence from courses and other program experiences they teach or facilitate to address research questions in teacher education as well as in their fields within teacher education.
- 2) Build a useful and usable structure for collecting evidence to assist faculty who seek both extramural and intramural support for their scholarly work around teacher preparation.
- 3) Aid the Division of Teaching and Learning to address multiple standards for accreditation purposes in a substantive way. These include: (a) NCATE Standard 1 (Candidate Knowledge, Skills, and Professional Dispositions); (b) NCATE

Standard 2 (Assessment System and Unit Evaluation) (c) Title 2 (Federal Law; Sections 205-208 of the Higher Education Act) and the (d) Educational Professional Standards Board (State board responsible for issuing and renewing certificates for teachers and administrators).

- 4) Model for candidates and practicing teachers, who are our students, a culture of inquiry and evidence that serves the needs of learners and contributes to the body of research relevant to our work as educators and scholars.

## **Study Design and Methods**

### ***Design***

The proposed work described here intends to create a line of research around Teaching and Learning courses and other program experiences and the outcomes or effects of those courses and program experiences. That is, researchers consider both the processes and outcomes of courses and other program experiences. By their nature, artifacts from Teaching and Learning courses and program experiences present opportunities for quantitative, qualitative, and mixed methods research (Creswell, 2009). In other words, assignments and other activities include, for example, written assignments submitted electronically, hard-copy papers and reflection tasks, and multiple-choice assessments of and on candidate perspectives and growth.

Additionally, the continuous assessment system used across the division provides ongoing, systematic processes for collecting and analyzing evidence. The ongoing assessment system is routinely used for Student Learning Outcome reports, analyses of candidates' performance on particular assignments and on specific candidate's performance across assessments.

### ***Threats***

As practitioner-researchers, faculty are cognizant of the constraints and limitations to their work. These include cautions with respect to backyard research as well as methodological and epistemological concerns (Glesne, 1999).

### ***Research Questions***

This study is designed with four broad, primary research questions. Additionally, Attachment A includes a table listing examples of field-specific research questions aligned to these as well as examples of relevant data analyses employed to answer those specific questions:

- (1) To what extent (how) do courses and other program experiences (curriculum, instruction, assessments) influence candidate learning in the division of Teaching and Learning?
- (2) How do specific courses and instructional practices influence candidates?

(3) What are candidates' perceptions of experiences in Teaching and Learning programs?

(4) How do candidates perform across and within programs?

### ***Data and Assessment Sources***

Demographic data and information typically used for instructional purposes (e.g. certification areas) may be used. Any data which may identify individual students (such as names and student identification numbers) are confidential, and all procedures used by instructors to maintain confidentiality are applicable.

As mentioned, the division uses a continuous assessment system. This system is called CARDS. The technology-based CARDS (Continuous Assessment Review System) model uses LiveText®, PeopleSoft®, and iStrategy®. The CARDS System was developed with the assistance of UofL's Institutional Research and Information Technology. It has attracted national inquiries and has been promoted by the LiveText® and iStrategy® companies. Data from the CARDS system are collected using LiveText®. Candidates complete course Hallmark Assessment Tasks (HATs) and portfolios in LiveText®, and faculty (including clinical faculty and university supervisors) use standards-based electronic assessments within LiveText® to assess the work submitted. Through the use of the electronic assessments, faculty and administrators have access to reporting capabilities so that overall student performance can be monitored at the course level as well as at the program level in relationship to national, state, and unit standards.

Data from CARDS is used for several purposes. Some of the data captured in LiveText® are used to populate candidate transition points in PeopleSoft® iStrategy® is then used to provide reports on the transitions of candidates and aggregate data for programs. The reports are used by program faculty for completing university Student Learning Outcomes reports (SLOs), for determining program improvements, and for internal and external accountability needs. The use of CARDS data for this empirical study is an extension of these uses.

Additional course and other program experience artifacts maybe used as evidence in self-study research as well. During class meetings, candidates participate in a variety of writing assignments and other tasks as part of normal instructional practices. Before and after class meetings, they participate in asynchronous discussions via Blackboard and e-mail with peers and instructors. Students complete assessments in Livetext and Blackboard as part of their own self-evaluations as well as to provide formative and summative assessments of their learning for instructors.

Course and program experience evaluations include several items that identify instructional materials and methods for student evaluation. Instructors and facilitators may choose to use their own evaluations as sources of data.

Attachment B includes a list of course artifacts and relevant ancillary items referenced above.

### ***Analyses***

Performance on assessments are typically reported numerically and as performance criteria (e.g. Target, Acceptable, Not Acceptable). Thus, numerical data may be used to report descriptive statistics. Other quantitative measures, including the use of inferential statistics, will depend on the sample size and applicability of the data used.

Qualitative analyses will apply when instructors require more in-depth analysis of evidence, particularly when context and process are important.

### **Inclusion and Exclusion Criteria**

All current Teaching and Learning courses and program experiences from (a) Special Education; (b) Elementary and Early Childhood Education; (c) Middle and Secondary Education; and (d) Physical Education will be included. Additionally, all course artifacts from those taught in previous semesters will be included.

### **Human Subjects Protections**

Candidates receive the following notice in all course syllabi and purchase Livetext © as part of their program requirements:

All students enrolled in College of Education and Human Development (CEHD) programs are required to have a LiveText© account. LiveText© will be utilized for submitting a Hallmark Assessment Task (HAT) in every course in addition to other requirements by program (i.e., portfolios). If students do not already have a LiveText© account, they will be required to purchase one for use during the courses in which they are currently enrolled.

A LiveText© student membership may be purchased at [www.livetext.com](http://www.livetext.com) or from the University bookstore. The student membership fee is \$98 for five years. Additional years may be added to an account at a reduced rate for those students who remain at the University of Louisville beyond the life of the five-year subscription.

As mentioned, LiveText© is used across the College for the analysis and reporting of student learning outcome data (SLO's) to the University as well as for the analysis and reporting of student assessment data for accreditation.

Data storage protections included in LiveText© include multiple layers of protection for user data. The application contains a web-server farm that sits behind the first set of firewalls and makes requests of protected database and file storage services, which are protected by usernames, passwords and access-control lists. Intrusion detection and prevention services also run internal to the network. Students can only submit assignments under their own password-protected account, and only the course instructor and technical support members may access the assignments once submitted.

Candidates will receive information concerning the conduct of this study as part of their course information. Instructors will distribute information about the conduct of self-study in the division as an informational page attached to their course syllabi as well as a Notification of non-participation form that students complete and return to a third party. This form explains to candidates that they may designate that any course and other program experience artifacts that can be linked to them because his/her name is attached or otherwise labeled may not be used for research purposes.

It is not expected that participants would receive any benefit from this study, other than the professional growth that they have already experienced through completing the course assignment or program experience. Because identifiers are being removed, the potential for risk is similarly extremely low.

Every semester during a Division Wide faculty meeting, the primary investigator, Dr. Shelley Thomas will distribute information to all faculty members about the self study project. This information will include the following:

- IRB number and copy of the approved protocol
- Checklist of the Human Subjects Protections for the Self Study  
Paragraph for syllabi  
Notification of non-participation form for students

## References

- Carter, K. (1990). Teachers' knowledge and learning to teach. In W. R. Houston, M. Haberman, & J. Sikula (Eds.), *Handbook of research on teacher education* (1st ed., pp. 391-410). New York: Macmillan.
- Cochran-Smith, M., & Zeichner, K. M. (Eds.) (2005). *Studying teacher education: The report of the AERA Panel on Research and Teacher Education*. Mahwah, N.J.: Lawrence Erlbaum Associates.
- Cochran-Smith, M., et al. (2009). "Re-Culturing teacher education: Inquiry, evidence, and action. *Journal of Teacher Education*, 60(5), 458-468.
- Creswell, J.W. (2009). *Research design: Qualitative, quantitative, and mixed methods approaches*. Thousand Oaks, C.A.: Sage Publications, Inc.
- Feiman-Nemser, S. (2001). From preparation to practice: Designing a continuum to strengthen and sustain teaching. *Teachers College Record*, 103(6), 1013-1055.
- Schwartz, H. (1996). The changing nature of teacher education. In J. Sikula (Ed.), *Handbook of research on teacher education* (2nd ed., pp. 3-13). New York: Macmillan.



**Attachment A**

| <b>Primary Question</b> | <b>Examples of Field-Specific Questions</b>   | <b>Examples Evidence of Candidate Learning also used as Empirical Data</b>  | <b>Analysis</b>             |
|-------------------------|---|---|-----------------------------|
| 2, 4                    | What is the impact of using technology on the reflective practices of pre service teachers enrolled in Teaching and Learning Courses? | <b>Performance on Assessments In-Class and Out-of-class Assignments:</b><br>Online blogs completed in Blackboard, course quizzes and exams completed in Blackboard, text-based reflective notebooks | Quantitative<br>Qualitative |
| 2, 4                    | What is the impact of using technology on the content knowledge of candidates enrolled in Teaching and Learning Courses?              | <b>Performance on Assessments In-Class and Out-of class Assignments:</b><br>Online blogs completed in Blackboard, course quizzes and exams completed in Blackboard, text-based reflective notebooks | Quantitative<br>Qualitative |
| 1                       | How do candidates understand and acquire specific teaching skills?  | <b>Performance on</b>   | Quantitative<br>Qualitative |

|   |   |   |                             |
|---|---|---|-----------------------------|
|   |   | <b>Assessments In-Class and Out-of class Assignments:</b><br>case studies,<br>lesson plans,<br>papers,<br>reflections,<br>text sets                                       |                             |
| 2 | Were course objectives (for social studies methods course) met as evidenced by student performance on course materials? | <b>Performance on Assessments In-Class and Out-of class Assignments:</b><br>lesson plans,<br>I-search,  | Quantitative<br>Qualitative |
| 4 | How do candidates exhibit critical thinking skills?   | <b>Performance on Assessments; In-Class and Out-of class Assignments:</b><br>Critical Incident Reports,<br>Teacher Work Samples   | Quantitative<br>Qualitative |
| 3 | How are candidate's perceptions of diversity influenced by course experiences?  | <b>Performance on Assessments; In-Class and Out-of class Assignments:</b><br>Double entry journals,<br>reflective papers,<br>gallery walks,<br>surveys (self-assessments) | Quantitative<br>Qualitative |

Attachment B

Relevant Ancillary Items

Table 12:Initial Certification Programs: Continuous Assessment Record and Documentation System (CARDS 1-3)

Table 13:Advanced Certification Programs: Continuous Assessment Record and Documentation System (CARDS 4-6)

National Council for Accreditation of Teacher Education (2008). Professional Standards: Accreditation of Teacher Preparation Institutions  
Standard 1: pp. 16-24  
Standard 2: pp 25-29  
Standard 4: pp. 34-37

## APPENDIX E

### Addendum to IRB Study

Inquiry and Evidence: Self Study across the Division of Teaching and Learning,  
College of Education and Human Development

#### Additional Objectives:

- Identify strengths and weaknesses of the CEHD assessment system to promote revision of the system to support continuous improvement of educator preparation programs.
  - To determine the reliability of instruments used to measure student knowledge, skills, and dispositions in the CEHD assessment system with the intent to improve system quality and accuracy.
  - To examine construct validity of assessment items aligned with the conceptual framework constructs of Inquiry, Action, and Advocacy to better understand the alignment of constructs to current assessments.
  - To identify the impact of data from the CEHD assessment system on program improvement decisions and processes.

#### Additional Modifications:

- Katie Shanahan is a staff member engaged in the implementation and development of the CEHD assessment system as well as doctoral candidate in the Department of Educational Leadership, Foundations, and Human Resource Education.
- Katie will be reviewing both initial and advanced educator preparation program data and SLO reports in the CEHD.

## CURRICULUM VITA

NAME: Katherine Amelia Shanahan

ADDRESS: College of Education and Human Development  
South 1<sup>st</sup> Street  
University of Louisville  
Louisville, KY 40292

EDUCATION & TRAINING: B.A., Computer Science  
DePauw University  
1999-2003

M.Ed., Instructional Technology  
University of Louisville  
2006-2009

Ph.D., Educational Leadership and Organizational  
Development, Evaluation  
University of Louisville  
2009-2014

PROFESSIONAL EXPERIENCE: ***Instructor***  
Department of Leadership, Foundations and Human  
Resource Education  
College of Education and Human Development  
University of Louisville, Louisville, KY  
Fall 2013

***Assessment Coordinator***  
*Office of the Dean*  
College of Education and Human Development  
University of Louisville, Louisville, KY  
2005 – Current

***General Education Assessment Project  
Technology Coordinator and Assessment Support  
Specialist***  
University of Louisville, Louisville, KY  
2010 – Current

***Academic Technology Support Specialist***

Indiana State University, Terre Haute, IN  
2004-2005

PROFESSIONAL EXPERIENCE: ***Intermediate Statistics***

Department of Leadership, Foundations and Human  
Resource Education  
College of Education and Human Development  
University of Louisville, Louisville, KY  
Fall 2013

***Tutor, Applied Statistics***

5 students in Education Administration, EdD  
College of Education and Human Development  
University of Louisville, Louisville, KY  
Summer 2013

GRANT EXPERIENCE:

***Grant Evaluator, Collaborative Teacher Training  
in Content Literacy***

College of Education and Human Development  
University of Louisville, Louisville, KY  
2012-2013

***i2a Unit Implementation Grant Assessment  
Consultant***

College of Education and Human Development  
University of Louisville, Louisville, KY  
2010-2011

COMMITTEE SERVICE AND  
COMMUNITY ENGAGEMENT  
EXPERIENCE:

***Technology Committee (Ad-hoc member)***, College of Education and Human  
Development, University of Louisville, Louisville, KY,  
2011-Present

***Kentucky Information Technology for Educator Preparation (K-ITEP)***.  
State-wide collaborative to support information technology integration  
and system development at the state and institutional level, 2005-Present

***College Educator Preparation Committee***, College of Education and Human  
Development, 2007-Present

**University Educator Preparation Committee**, College of Education and Human Development, 2007-Present

**CEHD Dean's Search Committee** (Staff Representative), 2011

**Consultant, LiveText™** (technology solutions for strategic planning, assessment, and institutional effectiveness), provided guidance and feedback during the development, design, and pilot stages of the C1 version of LiveText and Field Experience Module), 2006-Present

#### NATIONAL MEETING PRESENTATIONS:

***Critical Thinking as a Powerful Practice in Teacher Education.*** Selected Roundtable. Track 4: Partnership in Education. 19<sup>th</sup> Annual Conference of Coalition of Metropolitan and Metro Universities. Transforming and Sustaining Communities through Partnerships. Louisville, KY. October 2013 (with Ann Larson and Joanne Webb).

***Kentucky PK-20 Data Collaborative: Data Repository Development.*** Selected Session. 65<sup>th</sup> Annual Meeting of the American Association of Colleges for Teacher Education, San Francisco, CA. March 2013 (with Ann Larson, Gary Schroeder, Linda Nickel, Manish Sharma, and Margaret Moore).

***Assessing Key Candidate Development. Critical Thinking as Powerful Practice: Revising Assessments in Educator Preparation as the Next Step Toward Excellence.*** Selected Session. 64<sup>th</sup> Annual Meeting of the American Association of Colleges for Teacher Education (AACTE). Chicago, IL. February 2012 (with Ann Larson, Joanne Webb, and Cheryl Kolander).

***Writing the Next Chapter: Strengthening a Continuous Assessment System through Increased Alignment with University Accountability Structures.*** Selected Session. 63<sup>rd</sup> Annual Meeting of the American Association of Colleges for Teacher Education (AACTE). San Diego, CA. February 2011 (with Ann Larson, Cheryl Kolander, and Joanne Webb).

***Creating a Collaborative and Innovative Data Model: One University's Journey toward Assessment, Accountability, and Accreditation.*** Selected Session. Annual Conference Southern Association of Colleges and Schools. Louisville, KY. November 2010 (with Ann Larson, Joanne Webb, and Cheryl Kolander).

***Creating a Collaborative and Innovative Data Model: One University's Journey toward Assessment, Accountability, and Accreditation.*** Selected Session. 62<sup>nd</sup> Annual Meeting of the American Association of Colleges for Teacher Education. Atlanta, GA. February 2010 (with Ann Larson, Cheryl Kolander, Robert Goldstein, and Joanne Webb).

***Institutional Assessment Initiatives and the Educator Preparation Unit: Creating a Collaborative and Innovative Data Model.*** Selected Session. 62<sup>nd</sup> Annual Meeting of the American Association of Colleges for Teacher Education (AACTE). Atlanta, GA. February 2010 (with Ann Larson and Cheryl Kolander, University of Louisville; Russ Wall, Alesa Walker, and Brenda Nix, Murray State University; Caroline Everington and Jennie Rakestraw, Winthrop University; Caroly Ryan, Northern Kentucky University; Thillainatarajan Sivakumaran, Christy M. Riser, and Dorothy Schween, University of Louisiana at Monroe; Gary Schroeder, University of Kentucky).

***Student Retention and Success: Using the Data System to Admit and Retain a Diverse Student Body.*** Selected Session. 62<sup>nd</sup> Annual Meeting of the American Association of Colleges for Teacher Education (AACTE). Atlanta, GA. February 2010 (with Gary Schroeder and Rosetta Sandidge, University of Kentucky; Ann Larson and Cheryl Kolander, University of Louisville; Alesa Walker, Russ Wall, and Brenda Nix, Murray State University).

***Creating a Collaborative and Innovative Data Model: One University's Journey toward Assessment, Accountability, and Accreditation.*** Selected Session. Annual Meeting of the National Evaluation Institute (NEI). Louisville, KY. October 2009 (with Ann Larson and Joanne Webb).

***One University's Journey Toward Assessment, Accountability, and Accreditation: Assessment and Accountability using PeopleSoft and iStrategy.*** Selected Session. Annual Meeting of the iStrategy Users. Timonium, MD. October 2009 (with Joanne Webb and Robert Goldstein).

***Unit-Wide Continuous Assessment, 10 Unit Key Assessments, Transition Points, and Student Learning Outcomes.*** Selected Session. LiveText Collaboration Conference. Chicago, IL. July 2009

***CEHD Continuous Assessment System.*** Selected Session. 61<sup>st</sup> Annual Meeting of the American Association of Colleges for Teacher Education (AACTE). Chicago, IL. February 2009 (with Ann Larson and Joanne Webb).

***Unit-Wide Continuous Assessment, 10 Unit Assessments for Initial Certification and Advanced Programs.*** Selected Session. LiveText Collaboration Conference. Chicago, IL. July 2008 (with Tricia Bronger).



***Utilizing LiveText for Self-Study, Data Collection and Manipulation, Program/Course Assessments, Electronic Requests for Field Experience and Clinical Placements.*** Selected Session. LiveText Collaboration Conference. Chicago, IL. July 2007 (with Tricia Bronger).

***Utilizing LiveText for Self-Study, Data Collection and Manipulation, Program/Course Assessments, Electronic Requests for Field Experience and Clinical Placements. Selected Session.*** LiveText Collaboration Conference. Chicago, IL. July 2006 (with Corrie Orthober).

***Stressing the Value of LiveText to a Student. Selected Session. LiveText Collaboration Conference.*** Chicago, IL. July 2006 (with Heather Millick).

***Best Learning Practices By COE Are Contagious, How Indiana State University Went University-Wide with LiveText.*** LiveText Collaboration Conference. Chicago, IL. July 2005

#### STATE AND REGIONAL MEETING PRESENTATIONS:

***Redefining Educator Preparation: Best Practices and Collaborations in Preparing Next Generation Teachers.*** Selected Session. Annual Conference of the Kentucky Association of Teacher Educators, Erlanger, KY. September 2013. (with Joanne Webb, Ann Larson, and Danna Morrison)

***Developing Critical Thinking Educators for the 21<sup>st</sup> Century Classroom.*** Selected Session. Annual Conference of the Kentucky Association of Teacher Educators (KATE). Georgetown, KY. September 2012 (with Ann Larson and Joanne Webb).

***Strengthening Critical Thinking in the Teacher Education Program. Selected Session.*** Annual Conference of the Kentucky Association of Teacher Educators (KATE). September 2011 (with Ann Larson and Joanne Webb).

***Creating a Collaborative and Innovative Data Model: One University's Journey toward Assessment, Accountability, and Accreditation.*** Selected Session. Annual Conference of Kentucky Association for Institutional Research (KAIR). Louisville, KY. October 2010 (with Ann Larson, Joanne Webb, and Robert Goldstein).

***Using Critical Thinking to Enhance 21st Century Educator Preparation Programs.*** Selected Session. Annual Conference of the Kentucky Association of Teacher Educators (KATE). Georgetown, KY. September 2010 (with Joanne Webb, Cheryl Kolander, and Ann Larson).

***Using Critical Thinking to Enhance 21st Century Educator Preparation Programs.*** Selected Session. Annual Conference of the Kentucky Association of Teacher Educators (KATE). Georgetown, KY. September 2010 (with Joanne Webb, Cheryl Kolander, and Ann Larson).

***Creating a Collaborative and Innovative Data Model: One University's Journey toward Assessment, Accountability, and Accreditation.*** Selected Session. Annual Meeting of the Southeastern Regional Association of Teacher Educators (SRATE). Louisville, KY. November 2009 (with Ann Larson, Joanne Webb, and Robert Goldstein).

***LiveText as a Technology Tool for Accreditation, Data Collection and Manipulation, Program/Course Assessments, Electronic Requests for Field Experience and Clinical Placements.*** Kentucky Information Technology for Educator Preparation (ITEP) Symposium. Lexington, KY. November 2005 (with Natalie Stiglitz and Corrie Orthober).

***Possibilities and Capabilities of Electronic Portfolios for Faculty and Students. University of Louisville Delphi Celebration of Teaching and Learning.*** Louisville, KY. October 2005 (with Sherri Brown).

***Faculty and Student Perceptions of the Implementation of Electronic Portfolios at Indiana State University.*** Indiana Higher Education Telecommunication System (IHETS) and Indiana Partnership for Statewide Education (IPSE) Conference. Indianapolis, IN. April 2005 (with Kenneth Janz and Christy Coleman).