An Analysis of the Concurrent Validity of the Interactive Computer Interview System as Defined by Student Achievement on the Kansas State Assessments

By

Todd Wheat

Submitted to the graduate program in the Department of Educational Leadership and Policy Studies and the Graduate Faculty of the University of Kansas in partial fulfillment of the requirements for the degree of Doctorate of Education

Dissertation Committee:

_____

Chairperson, Dr. Howard Ebmeier

_____

Dr. Argun Saatcioglu

_____

Dr. Thomas DeLuca

_____

Dr. Marc Mahlios

_____

Dr. Rick Ginsberg

Date Defended: April 24, 2013

The Dissertation Committee for Todd Wheat certifies that this is the approved version of the following dissertation:

An Analysis of the Concurrent Validity of the Interactive Computer Interview System as Defined by Student Achievement on the Kansas State Assessments

_____

Chairperson, Dr. Howard Ebmeier

Date approved: May 30, 2013

Abstract

This study's primary research question was to determine whether or not the Interactive Computer Interview System displays concurrent validity as defined by student achievement through a criterion-referenced test. The criterion-referenced test used for this analysis was the Kansas State Reading and Math Assessment.

In this study, 40 third-, fourth- and fifth-grade teachers throughout the Olathe School District in Olathe, KS were interviewed using the ICIS screening tool. Mean Kansas Assessment scores were then collected for each of the 40 teachers in the subjects of reading and math from 2010, 2011, and 2012. A Z Score calculation and subsequent residual score was determined to adjust for differing test lengths, difficulty, and the effect of socio-economic status as defined by free/reduced lunch percentage for each individual teacher's class. Finally, a correlation analysis was performed between the mean Kansas Assessment Results from 2010-2012 in both reading and math and that same teacher's total weighted average on the ICIS.

In addition to a correlation analysis between a teacher's mean score on the Kansas State Assessments over a three year period and their ICIS total weighted average score, an analysis was also performed using an administrator rating system. Each principal that supervises the staff used in this study was asked to give a rating of that teacher's overall teaching ability on a scale of 1-10. These ratings were then compared to both that same teacher's ICIS total weighted average and their three year mean reading and math scores.

No significant correlation was found between a teacher's total weighted average on the ICIS and that same teacher's residual reading (p=-0.17816) or residual math (p=-0.16327). There was also no significant correlation found between the administrator's rating and a teacher's total weighted average on the ICIS (p=0.2498). However, there was a statistically significant correlation at the .05 level between an administrator's rating and their residual reading (p=.3718) and at the .01 level on residual math (p=.4309).

Acknowledgements

There are many family members, friends, and colleagues that helped make this study a reality. My dad for all of his support, both financially and emotionally. My brother for helping proofread not just this study, but countless other papers throughout my years in higher education. My mother-in-law for continuing to push me to keep working and not give up and my father-in-law for his help conducting forty teacher interviews. My best friend Larry, whose passion for Excel came in handy many, many times. My dissertation chair, Dr. Howard Ebmeier, for his willingness to talk through the many challenges in creating a study like this from start to finish. The teachers and administrators throughout the Olathe School District for their willingness to provide the data needed to carry out this study. My two boys, Henry and Owen, for their patience as their Daddy spent countless hours away from home working on this study.

In particular, I want to thank my Mom who was able see me start my doctorate coursework, but wasn't able to make it to see the culmination of the passion and work ethic she instilled. I'm dedicating the completion of this program and degree to her memory.

Finally, I want to thank my wonderful wife, Ruth. The sacrifices she has made since this journey began are indescribable. I will never truly be able to express how grateful I am for all she has done for our family and me. She is an inspiration.

TABLE OF CONTENTS

List of Tables

List of Tables (cont.)

List of Graphs

I. Introduction

The expectations and responsibilities of school districts and ultimately teachers have clearly increased over the years with the inception of the *No Child Left Behind Act of 2001* (NCLB). It is apparent that the need for exceptional teachers is critical to programs aimed at improving student performance, and teacher quality has been cited numerous times as one of the most important factors contributing to student achievement. Additionally, the United States Department of Education estimates that schools will need to hire more than two million new teachers over the next decade (U.S. Department of Education, 2002), thus school systems across the nation are now faced, or will be faced, with the daunting task of searching through virtually thousands of applications to find and hire the most competent teachers. Currently millions of employment screening interviews are conducted each year by school districts to identify teachers that are the best fit for the school district and local campus (Reik, 2007). Sorting and selecting the highest quality teachers from a seemingly endless pool of candidates can be an arduous task for school administrators. In an effort to counter this, a number of school districts have opted to take advantage of various commercial screening tools such as the Gallup's Teacher Perceiver Interview/TeacherInsight, Kenexa's Star Teacher, and the American Association of School Personnel Administrator's (AASPA) Interactive Computer Interview System (ICIS). Nearly 2,000 districts within the United States are currently using some form of commercial screener (Delli, 2001). This equates to nearly 15 percent of the nation's school districts. The fundamental question school districts need to ask is whether or not these commercial

screening tools such as the TPI or ICIS are reliable and valid predictors of teacher job performance.

A number of studies have been conducted on commercial screening tools and their ability to accurately predict teacher effectiveness (Metzger and Wu, 2003; Allshouse, 2003; Evans, 2003; Weishaar, 2006; Riek, 2007). However, few studies have provided a comparative analysis on these screening instruments and their correlation to student achievement. This study sought out to determine if employment screening instruments commonly used in the field were correlated with academic achievement tests. The ICIS instrument was selected as a typical model of an employment interview instrument for use in this study for two reasons. First, the two historically most common instruments (Gallup and Haberman) are no longer commercially supported since they have recently migrated to an on-line format and can no longer be considered a dyadic interview. Second, only the ICIS instrument was available for use in this study since permission for use of the older (original) versions of the Gallup and Haberman instruments could not be obtained. Thus, the ICIS screening tool was selected as a representative of typical employment screening instruments. Therefore, the primary research question for this study was whether or not the ICIS displayed concurrent validity where the outcome variable was defined as student achievement on the Kansas State Reading and Math assessments.

The first phase and primary purpose of this study consisted of examining the correlation between a teacher's score on the ICIS employment interview instrument and the average score of that same teacher's students on the Kansas State Assessments. Forty teachers were interviewed by an unbiased retired principal trained on

appropriately administering the ICIS. The Kansas State Assessment scores for the students of those forty teachers were collected in reading and math for a three-year period and then correlated with the results of their ICIS interview score.

The secondary analysis consisted of comparing the ICIS interview data collected in the first phase of the study with principal ratings of those same teachers. Since prior research (Allshouse, 2003; Cowan, 1999; Smith, 2006) indicated a significant correlation existed between the two variables, the secondary analysis was undertaken in part as a replication study and to better understand the context of the experimental design, subject selection, measurement error, or other possible design flaws. In addition, since the administrators' rating of teacher had been gathered in the secondary phase of the study, it also facilitated an analysis that included correlating the administrators' rating with teacher's class achievement scores on the same Kansas State Assessments. The secondary research question for this study was to determine whether or not the ICIS screening tool had a correlation to an administrator's rating on overall teaching effectiveness. An examination of correlations between administrator ratings and student achievement was also performed.

Significance of the Study

The goal in any employment process, is to hire the most competent, skilled, and knowledgeable individual available for the position. Research shows that teachers are the most influential school-based resource in student learning and that good teaching does matter (Ferguson, 1991; Darling-Hammond, 2000; Wright, Horn, and Sanders, 2007). This is particularly important in a school setting, as school personnel are the

most costly part of any budget, and may remain in a single district for the course of their career, and have the greatest impact on the quality of the educational program provided (Smith & Ebmeier, 2009). Additionally, much of the research suggests the hiring of quality educators increases the likeliness of value being added to student learning (Bond et al., 2000; Cavaluzzo, 2004; Fisher & Dickenson, 2005). Pillsbury (2005) claims that the most important decision that principals make is hiring teachers. Clement (2009) proposed that, in today's era of accountability, we have high-stakes hiring, due to the potential for a weak hire to negatively impact student achievement, lower overall school performance, and lower morale of colleagues (p. 22).

The majority of organizational research on employment interviews over the past eighty years has highlighted the importance of structuring interviews in order to maximize their reliability as a sound decision-making tool (Delli & Vera, 2003, p. 138). With this concept in mind, one can infer that the selection process of hiring the best candidates should be carefully structured to provide the best outcome in terms of predictability of staff performance.

A number of studies admonish the employment interview as an effective tool as it is typically performed in the traditional sense (unstructured) (Emley & Ebmeier, 1997; Harvey & Struzziero, 2000). Thayer (1978) indicated several criticisms of the traditional interview include weaknesses in the way the information is gathered, judgment bias, and errors in decision making. Interviewers may inadvertently influence responses through nonverbal behavior, ask questions not adequately aligned to the essential responsibilities of the job, or lose control of the interview by talking too much or too little when follow-up questions would be helpful. Despite these limitations, however, the

interview is likely to remain a popular tool for employee selection (Carlson, Thayer, Mayfield, & Peterson, 1971; Murray, 1990; Huffcutt, Conway, Roth, & Stone, 2001; Reik, 2007; Doll 2009).

With the necessity and importance placed on the interview as part of the selection process, a significant amount of research has also been conducted on the reliability (Conway, Jako, & Goodman, 1995) and validity of the employment interview (Huffcutt & Arthur, 1994; McDaniel, Whetzel, Schmidt, & Mauer, 1994; Wiesner & Cronshaw, 1998; Wright, Lichtenfels, & Pursell, 1989). In general, these studies do suggest that the employment interview has the potential to predict job performance, so long as the interview is structured (Huffcutt, Conway, Roth, & Stone, 2001). However, most of these studies have been conducted employing a rating of employees by their supervisors as the correlated variable. Only a handful of studies (Reik, 2007; Weishaar, 2007) have actually examined student achievement as the outcome variable. This study specifically focuses on teacher classroom effectiveness in terms of increasing students' academic test scores as the outcome variable and thus represents a significant variance from the bulk of extant studies.

Another consideration for this study was the determination of which data to use as the outcome variable of student achievement data. The Kansas State Assessments were selected due to the impact these assessments have on public schools throughout the state of Kansas. These scores not only determine whether or not a school has made Adequate Yearly Progress (AYP), but are also reported out to the media so the general public can make a decision on whether or not they feel their children are getting a quality public education. The implications of a school that does not make AYP include

being identified as a school “on improvement”, a publicized status on the Kansas State Department’s annual report card, and the potential of lost funding due to a family’s option to have their child attend a school that has not been identified as on improvement (Kansas Department of Education, 2012).

II. Review of Literature

This section of the study explored the evolution of the employment interview as a tool for hiring the highest quality employees. Research on how the employment interview has evolved over the past six decades is examined. Additionally, three screening tools specifically intended to be used in education are described: Gallup's Teacher Perceiver Interview and TeacherInsight, Haberman's Star Teacher, and AASPA's Interactive Computer Interview System. Finally, the relationship between principal evaluations and teacher effectiveness is explored to address the secondary analysis of this study.

History of the Employment Interview for Teacher Selection

The hiring process has always identified the employment interview as an integral component (Eder & Ferris, 1989). Foundational research dating back to as early as 1952 established the employment interview as playing a substantial role in the hiring process (Asch, 1952). Throughout time, the personal interview has been the foundation and most common tool utilized within the hiring process in education and other fields (Castettern and Young, 2000; Eder, 1989). Because of the popularity of and importance placed upon the interview in the selection process, much research has been conducted over the past few decades to investigate the effectiveness of the employment interview. There have been a number of meta-analyses of interview validity (Huffcutt and Arthur, 1994; McDaniel, Whetzel, Schmit, and Maurer, 1994; Wiesner and Cronshawk, 1988; Wright, Lichtenfels, and Pursell, 1989) and of interview

reliability (Conway, Jako, and Goodman, 1995). Taken as a whole, these studies suggest that employment interviews do have the potential to help predict job performance (Huffcutt, Conway, Roth, and Stone, 2001) when the interviews are structured. Methods to increase structure include note-taking during the interview, asking all applicants the same questions based on skills and requirements of the job, having clear evaluation criteria, and using a consistent approach to administration with well-trained interviewers.

A few foundational studies have suggested that the typical employment interview often has low reliability and validity (Wagner, 1949; Mayfield, 1964). A study conducted by Baskin, Ross, and Smith (1996) concluded that most professions require understanding beyond simple basic skills and truly successful hiring practices come down to determining the personality traits and abilities required for effective performance (Baskin, et al, 1996.).

Teacher credentials, experience, and references all contribute to the selection of high quality teachers, but the face-to-face interview is likely to be the most frequently used method of teacher selection and quite possibly carries the most weight (Whetzel, Baranowski, Petro, Curtin, & Fisher, 2003). Employment interviews are generally either structured, with a fixed set of questions that are scored against behaviorally based rating scales, or unstructured, where there is no standard protocol for questions or scoring (Whetzel et al., 2003).

Typically, the unstructured interview tends have a good deal of variance from one interviewer to another. Antoline (2000) concluded that unstructured interviews often result in question variation, ambiguous definitions of what constitutes as quality

responses, and eventually diminished reliability. Finding ways to add structure to an employment interview system does appear to increase the level of validity and reliability (Conway, Jako, and Goodman, 1995). Additionally, research has shown the presence of structure in an interview can positively impact the issue of validity due to the improved emphasis on areas such as social skills, organizational fit, applied mental skills, and overall job knowledge (Huffcutt, Conway, Roth, & Stone, 2001). In a structured interview, a standard set of questions would be posed to all candidates, allowing the responses to be compared between all candidates (Reik, 2007).

To further define the difference between structured and unstructured interviews, Huffcutt (1992) depicted four levels of interview structure: Level 1 had no guidelines for questions and no benchmarks with which to evaluate questions. Level 2 specified question topics and moderate evaluation of responses. Level 3 specifically provided questions and a definite rating scale used to evaluate responses. Level 4 required exact questions asked of every candidate and responses were evaluated based upon specified benchmark answers. Traditionally school officials used a combination of Levels 2, 3, and 4 to interview potential teachers. Goldstein (1986) emphasized that the interaction and an appropriate level of questioning between individuals during interviews was necessary in order to ensure there was a good fit. He further stated that the level of questioning should address the candidate's ability to analyze a problem, to organize different facets of a problem, to explain the conditions, and to mitigate or solve the problem (Goldstein, 1986). According to Whetzel et al. (2003), interview questions were often intended to determine problem-solving abilities through two different question types: behavioral and situational. Behavioral questions required candidates to consider

an instance in which they had to deal with a specific situation, and situational questions generally involved a scenario for which candidates were asked how they would handle the circumstances (Whetzel, Baranowski, Petro, Curtin, & Fisher, 2003). Both behavioral and situational questions have been tested and found to be valid (Taylor & Small, 2002).

The aforementioned research has provided a framework for the expansion of the structured interview (Baker and Morris, 1984). The Teacher Perceiver Interview (now TeacherInsight) (Gallup, 1997) and the Star Teacher Interview (Haberman, 1993) were two structured interview tools frequently used by school systems. Studies on the predictive validity of these two interview tools have been met with varying levels of success (Baskin, Ross, & Smith, 1996; Carney & Johnson, 1995; Delli & Young, 2002; Haberman, 1993; Metzger & Wu, 2003). The Interactive Computer Interview System (ICIS) is a third structured interview, also met with varying degrees of success (Smith, 2006; Reik, 2007; Weishaar, 2007) and will be employed as a representative dyadic employment interview of those typically found in use in the field.

Employment Interview Instruments Used in Education

To translate affective beliefs, attitudes, and values into practicable teacher selection, many school districts have turned to commercial teacher hiring instruments. After decades of research, three main organizations have emerged from the sea of commercial teacher selection instruments: The Gallup Organization's original Teacher Perceiver Interview, now transformed into an on-line survey called the TeacherInsight, The Haberman Foundation's Star Teacher with similar on-line transformation, and the

American Association of School Personnel Administrator's Interactive Computer Interview System. Some reviews of these and other commercial screening tools frequently used in education have been relatively negative citing the lack of internal reliability and failure to establish both content and predictive reliability (Metzger & Wu, 2003; Young & Delli, 2002; Baskin et al., 1996). Currently only the AASPA's ICIS still utilizes a face-to-face interview format. The other two screening tools discussed in this study only utilize an on-line version.

Gallup's Teacher Perceiver Interview and TeacherInsight

The Teacher Perceiver Interview was initially introduced in the 1970s by the Selection Research Inc. (SRI) and was widely considered the most frequently used commercial teacher interview (Delli & Young, 2002) until it was discontinued in 2006. The TPI was based on a belief that the personalities and potentialities a person brings to the field of education are essential elements in achieving excellence in teaching. Additionally, it was developed on the belief that these qualities can be measured through a structured interview process (Chalker, 1981). This structured interview tool was created around the concept of "life themes" and questions were developed to fit the themes, and answer keys that outlined key words were written (Chalker, 1981).

SRI acquired the Gallup Organization and adapted its company name in 1988 and later published TeacherInsight, another teacher selection tool to replace TPI. The difference between the original version of the TPI and the current TeacherInsight is that the TPI was a dyadic instrument, whereas TeahcerInsight is an on-line tool that utilizes a Likert scale based on 50 question multiple choice questions. Gallup claims that the

TeacherInsight tool can help the school hire the best teacher fast. On its webpage, it is stated:

> "TeacherInsight provides a quick, effective way to source and assess a large volume of applicants. This innovative selection tool allows you to:
>
> - Identify more teacher candidates like your best teachers
> - Increase the speed of assessing applicants
> - Reduce staff time spent interviewing applicants and the associated costs
> - Focus valuable district staff time on recruiting candidates
> - Keep you HR office open 24/7" (Gallup, 2009)

The Gallup Organization's webpage also claims that the TeacherInsight tool assesses the talents that result in teacher excellence that are difficult or nearly impossible to teach (Gallup, 2009).

> "The TeacherInsight assessment requires approximately 30 minutes to complete and is available 24 hours a day, 7 days a week. Results are based on the applicant's responses and include a score that is predictive of an applicant's potential for teaching success based on his or her talent. Once an applicant has completed TeacherInsight, access to the results is nearly immediate. Gallup instantly reports scores to districts through its Web-based reporting site, Gallup Online. Districts can also automate applicants accessing TeacherInsight and recording the results into existing applicant tracking systems." (Gallup, 2009)

The TeacherInsight assessment tool (The Gallup Organization, Princeton, NJ,http://education.gallup.com) is an electronic version of the Teacher Perceiver Instrument developed by the Gallup Organization. The TeacherInsight is composed of two parts, one being the interview portion and the other being the “StrengthsFinder” development tool.

The TeacherInsight interview requires approximately 40 minutes to complete. The entire process is completed on-line and is composed of selected questions from the Teacher Perceiver Interview that includes statements that use a 1-5 Likert scale with “strongly disagree” to “strongly agree” response options; multiple choice items that are intended to reveal applicants’ attitudes, beliefs, and behaviors; and open-ended questions that applicants answer in their own words. The report presented by Gallup is based on responses provided by the applicant and is a “best prediction of the applicant’s potential for teaching success based on the applicant’s talent (Gallup, 2002).”

Scores can be reported through the Gallup Website or electronically provided to school districts (Gallup, 2002). The Gallup Organization conducted research utilizing a pilot Web interview that was given to 180 teachers across the United States (Wallwey, 2002). The data were obtained from focus groups and was composed of teachers, administrators, and students. The responses demonstrated the knowledge, skills, attitudes, and behaviors that exemplify outstanding teachers. These interview results, along with 30 years of Gallup research on the TPI indicating that “successful people respond differently from less successful people,” (Wallwey, 2002) were used to develop the pilot Web interview.

Information gathered from the pilot interview performance was used to develop a field interview (Wallwey, 2002). An analysis performed by Gallup examined the relationship between scores on the field interview and principal and student ratings of performances. Fourth- through 12th-grade students rated 111 teachers, while principals rated 159 participating teachers. The results indicated that “those teachers scoring high on the interview were more likely to be outstanding performers than those teachers who scored low on the interview” (Wallwey, 2002).

One of the predominant issues with the TeacherInsight is the issue of validity (Reik, 2007). One important aspect of validity is alignment in which the interview instrument must align between derivation and content. An early study on the TPI, (the foundation for TeacherInsight), may best frame this issue of alignment. According to this study, the TPI was found to be a better predictor of teacher popularity, than teacher effectiveness (Miller et al, 2007).

There have been a number of other studies on the validity of the TPI (Delli, 2001; Delli & Young, 2002; Haefele, 1978; Metzger & Wu, 2003). These studies showed a relatively low correlation to both teaching effectiveness and/or teaching quality (Delli & Young, 2002; Metzger and Wu, 2003). A study conducted by Delli & Young (2002) found no relationship between scores on the TPI as compared to principal ratings within the same theme. However, there did appear to be a more significant relationship between decisions made pre-employment to employee outcomes on the long version (compared to the short version) of the TPI (Delli & Young, 2002).

Metzger and Wu (2003) were succinct in their criticisms. They concluded that the claims made about the validity of the TPI in the TPI technical report and the SRI

were not supported. Metzger and Wu went on to claim that users of the TPI should be mindful that their study indicated the TPI's failure to meet minimal requirements as it pertains to instrument validity (Metzger and Wu, 2003).

The conclusions drawn by Metzger and Wu (2003) are echoed by a number of studies that found little to no statistical significance between a candidate's performance on the TPI or TeacherInsight and a variety of dependent measures. Aarestad (1980) found no relationship between the 34 first- through sixth- grade teachers interviewed using the TPI and how those students scored on math achievement tests. Gillies (1988) found a low correlation (0.17) between the 196 teachers that were interviewed using the TPI and how those same teachers were rated by each of their principals. Novotny (2009) examined the relationship between TeacherInsight scores and teacher performance as rated by an appraisal system known as the Professional Development Appraisal System (PDAS). This study concluded that only one of the eight domains which are part of the PDAS appraisal system correlated statistically significant (r=0.14).

Haberman's Star Teacher Interview

The Star Teacher Interview, created by Dr. Martin Haberman of the University of Wisconsin-Milwaukee, is sometimes referred to as the Urban Teacher Selection Interview. Haberman suggested there were two primary types of urban teachers: "Stars" and "Failures" (Haberman, 1995). Haberman (1995) states:

> Selection is significantly more important than training. It is easier and wiser to select people with attributes that will enable them to succeed in metropolitan schools than it is to expect that individuals who might be sexist, racist,

> uncreative, uninterested in the world of ideas, rigid, moralistic, humorless, or fearful will be transformed by virtue of completing a teacher education program (p.2).

According to the Haberman Educational Foundation, Inc., (2009) the Star Teacher Interview has been the culmination of over 30 years of research and development. Seven midrange functions (Table 2.1) were identified from what Haberman believed created a successful teacher. Haberman's seven midrange functions were influenced by the work of Columbia University sociologist, Robert K. Merton (Haberman, 1995).

Table 2.1: Star Teacher Interview Seven Midrange Functions

Function	**Brief Definition**
Persistence	Ability to work with challenging students.
Protecting Student's Learning	Student's learning is highest priority.
Application of Generalizations	Can apply principles and put them into action.
Approach to At-Risk Students	Ability to care for students of all backgrounds.
Personal/Professional Orientation	Balance personal feelings with professional responsibilities.
Burnout	Ability to function well in a bureaucratic institution.
Fallibility	How the teacher plans to deal with mistakes.

When giving the Star Teacher Selection Interview, the use of mid-range functions is used to predict future success of urban teachers. To predict teacher success, Merton identified traits and behaviors that he felt made teachers successful when teaching in urban settings and placed the traits on opposite extremes of a continuum (Table 2.2).

Table 2.2: Basis for Predicting Teacher Success (Haberman, 1995)

Personality Traits	**Mid-Range Function**	**Situational Demands**
What are the effective and constant teacher characteristics?		What behaviors would be effective for all teachers in a given situation?

The first extreme on the left side of this continuum represents personality traits that a person would exhibit in any situation. The right side represents behaviors a person would exhibit in a certain situation. A person operating at either extreme would be considered “dysfunctional. The basic principle of this model is that it is impossible to predict how a person would respond to a particular situation based exclusively on his or her personality traits (Haberman, 1995). As a result, Merton developed mid-range functions or behaviors that an individual would demonstrate in order to be effective.

Based on Merton’s research of mid-range functions, Haberman (1995) observed 124 student teachers in New York City who were deemed successful in an urban school setting. Haberman then began identifying which mid-range function they exhibited. Haberman’s first step was to determine which teachers would be considered “Stars” and which ones would be considered “Failures”. While comparing the extremes of the 124 individuals, Haberman formed the first mid-range functions for urban teachers. The original mid-range functions were eventually modified and refined to better align with the terminology with Haberman’s teacher interview (Table 2.3).

Table 2.3: Original and Current Terminology for Mid-Range Functions (Haberman, 1995)

Original Terminology	**Current Terminology**
Creativity, Problem Solving	Persistence
Human Relations Skills	Protecting Student's Learning
Planning	Application of Generalizations
Discipline	Approach to At-Risk Students
Teaming	Personal/Professional Orientation
Self-Analysis	Burnout/Fallibility

The first mid-range function is persistence. Haberman originally described persistence as creativity and problem solving. The questions form the Star Teacher Interview on persistence look for commitment, determination, and perception of the day-to-day functions of the job (Haberman, 1995). "Star" teachers were designated by being able to explore creative methods to solve problems that face urban schools. When presented a problem, teachers considered "Stars" would first define the problem, then evaluate the situation, and finally consider all options before making a final decision. These were teachers who were persistent had a great sense of self-efficacy. "Stars" repeatedly sought out solutions even when they felt the problem was perhaps irreversible.

The second mid-range function, protecting the learning of students, is also known as respect to authority. Originally defined as human relation skills, Haberman began to notice that even "Failures" were able to get along with colleagues and be liked in certain situations. A "Star" teacher that is protecting the learning of students is motivated to protect the learning environment of his or her students, even if that means questioning

or challenging school rules and norms. Haberman believed that “Stars” would avoid direct confrontation with administrators. As an alternative, they would simply win the administrator over by convincing him or her that the learning the students are experiencing is truly beneficial and therefore should be continued despite the expectations established by the school (Doll, 2009).

The third mid-range function was originally named planning, but later evolved into application of generalizations. The generalization is much broader than merely establishing a plan for teaching. “Star” teachers should be able to take principles and put them into action. Additionally, a “Star” will have strong follow-through skills, focused on what is best for teaching and learning. Haberman believed it was one thing for a teacher to state they believed in something or make a generalization, but if a teacher could cite *how* they put a generalization into practice, it would demonstrate that prospective teacher could move from general to specific (Doll, 2009).

The fourth mid-range function pertains to at-risk students. Originally identified as discipline, this particular mid-range function examines the accountability of a teacher when he or she likes (or dislikes) an at-risk student and evaluates them accordingly. The candidate should seek multiple ways to instruct a child despite his or her background. The principles and theories of child and adolescent development are typically based on middle-class values so everyone outside this range would be considered atypical (Haberman, 1995). To reject these theories would compel teachers to see themselves as inadequate. Simply blaming the student is much easier than questioning the whole foundation on which educational programs are built. A “Star”

teacher would not blame the student for their at-risk circumstances, but instead would take accountability for the students' learning no matter what (Doll, 2009).

The fifth mid-range function is personal versus professional orientation. "Star" teachers realize that although they will not like every student, they do have a professional obligation to accept responsibility for that student's learning. "Star" teachers are able to overcome the simple personal gratification of teaching and instead view teaching through a professional orientation. This type of teacher would be more likely to prosper in an urban setting and not let his or her personal needs interfere with the learning of a child (Doll, 2009).

The wording and terminology for the sixth mid-range function has been changed more than once by Haberman. Initially called teaming, this mid-range function progressed to bureaucracy and finally evolved to burnout. Large urban school districts are frequently caught up in a bureaucratic system which impedes the educational progress of the students served by the system (Haberman, 1995). "Star" teachers are aware of the ways the bureaucracy can exploit their efforts, but spend little time actually fighting it. Instead, "Stars" will accept the reality of the situation and find ways around the bureaucracy of urban schools without becoming cynical. "Star" teachers utilize their colleagues and resources to counter the issue of bureaucracy and burnout (Doll, 2009).

The final mid-range function is fallibility. Originally deemed self-analysis, this mi-range function was re-named because "Stars" do not focus only on themselves, but will accept others' mistakes and differences. According to Haberman, fallibility refers to a teacher's ability to reflect on his or her own mistakes and behaviors and begin to use and accept those actions of others. "Stars" understand mistakes are bound to happen,

and when they do, they can accept the mistake and learn from it. Conversely, “Failures” do not want to be held accountable for the mistakes they make and will place the blame on a colleague or on one of their students (Doll, 2009).

The seven mid-range functions presented by Haberman as required traits of an effective urban teacher are not in total congruence with much of the research on effective urban teachers. The Star Interview model claims teacher burnout is a critical factor for identifying successful urban teachers. However, in more recent studies, teacher burnout is not part of most of the research on effective urban teachers (Best, 2005; Klussman, 2004).

The original Star Teacher Interview was administered by asking each candidate two questions for each mid-range function, for a total of fourteen questions. After an initial response to each question, the interviewer would ask follow up questions for clarity and reliability. The interviewer would rate each response on a predetermined scale (Doll, 2009).

Today’s on-line version of the Star Teacher Interview involves 50 multiple choice questions, with 45 seconds provided to answer each question (The Haberman Educational Foundation, 2009). A summary sheet including the quartile in which the candidate scored is generated and able to be printed off (Metzger & Wu, 2003). This online version is available to the public and is different from the 15-20 question dyadic interview upon which the early research was based.

A relatively minor number of studies have been performed on the original Star Teacher Interview (Baskin, Ross, & Smith, 1996; Carney & Johnson, 1995; Haberman, 1993; Metzger & Wu, 2003). In 1993, Haberman conducted his own field study known

as the Milwaukee Trials which took place in the Milwaukee Public School system. The purpose of Haberman's study was to test the predictive validity of the Star Teacher Interview. In the Milwaukee Trials study, two independent groups of alternative-certification teacher candidates were examined. A candidate's interview ranking on the Star Teacher Interview was correlated to a performance ranking provided by their principals (Metzger & Wu, 2003). Haberman claimed a moderately strong relationship with both groups (r=.87 and r=.79) (Metzger & Wu, 2003). It should be noted that the sample size for both of these groups consisted of a very small 19 teachers.

Baskin et al. (1996) concluded Haberman's Star Teacher Interview was able to accurately predict student teaching success. A study conducted by Carney and Johnson (1995) examined principal ratings of first year teachers in the Minneapolis Public school system which were selected by utilizing the Star Teacher Interview and compared them to teachers selected using the more traditional means already in place. The study concluded that the teachers selected using the Star Teacher Interview were as good as (or better) than those selected using the standard hiring practices previously in place (Carney and Johnson 1995).

A study conducted by Klussman (2004), compared teacher performance on the current version of the Star Teacher Interview tool to student achievement as defined by reading and math student achievement scores. In this study, 87 reading teachers and 88 math teachers were interviewed in this study and the results found no statistical significance between the teacher's interview scores and student achievement on the 1351 student scores analyzed (Klussman, 2004).

AASPA's Interactive Computer Interview System

The third organization that has developed a teacher selection tool is the American Association of School Personnel Administrator's (AASPA) Interactive Computer Interview System (ICIS). The ICIS is a computer based interviewing system developed by Dr. Howard Ebmeier (2001) designed to be used in the hiring of Kindergarten-12th grade teachers (Ebmeier, 2001). The ICIS first tracks response patterns from the candidate and then generates questions based on those patterns. A detailed report is ultimately created based on the quality of responses provided by the candidate (Ebmeier, 2001).

To initiate an interview using the ICIS, the interviewer begins by submitting a password. The interviewer is then given an option of administering the short, normal, or long version of the program. Once the preferred version is selected, the interviewer poses the questions generated by the program and selects the response on the rubric which most closely aligns to candidate's response (Ebmeier, 2001). Table 2.4 details the possible number of questions for all three of the versions of the ICIS.

Table 2.4: Interview Instrument Question Allocation

Scale	**Short Version**		**Normal Version**		**Long Version**	
	Minimum Questions	Maximum Questions	Minimum Questions	Maximum Questions	Minimum Questions	Maximum Questions
Working with Others	3	5	4	6	6	8
Knowledge of Content	3	5	4	6	6	8
Knowledge of Teacher	6	10	8	14	12	20
Knowledge of Students	3	5	4	6	6	8
Total	**15**	**25**	**20**	**32**	**30**	**44**

Two primary resources were foundational in the development and selection of questions utilized in the ICIS; *Teacher of the Future and Praxis III: Classroom Performance Assessments*. The first of these two publications, *Teacher of the Future*, stemmed from the efforts of a national commission of school personnel officers (Ebmeier, 2001). This national commission collaboratively ascertained 11 skills and nine areas of knowledge which were believed to be essential characteristics for all teachers to possess (Ebmeier, 2001). Table A1 in the Appendix displays the 11 skills and nine areas of knowledge described in *Teacher of the Future*.

The second resource used to develop ICIS questions was the *Praxis III: Classroom Performance Assessments* (1995). A group of practicing teachers, in conjunction with Educational Testing Services, is credited with the creation of the Praxis assessment which is a requirement for beginning teachers to obtain appropriate licensure (Ebmeier, 2001). The Praxis III consists of 19 assessment criteria, organized into four separate domains (Table A2). The ten years of research put into the development of the Praxis III, was under the leadership of a national advisory committee (Ebmeier, 2006).

The questions for the ICIS were created using the documents referenced above and eventually divided into four sub-sections (Table A3). Criteria pertaining to what constitutes effective versus ineffective practice was based predominantly on the process-product research examined over the last three decades (Ebmeier, 2006). Scoring rubrics which accompany each question were later developed to add consistency. The interviewer is asked to score each response on a scale of Level 1 to

Level 3 (Level 1- ineffective; Level 2- effective; Level 3- highly effective) based on the quality of response as detailed in the rubric.

A large number of studies have been conducted on the ICIS (Allshouse, 2003; Evans, 2003; Cowan, 1999; Weishaar, 2007; Cox, 2006; Dillon, 2006; Dugan, 2007; Green, 2005; Longnecker, 2006; Smith, 2006). Specifically, a variety of studies have been conducted on either the construct or criterion related validity of the ICIS (Ebmeier, 2006).

Two studies specifically cited in the technical manual as having strong construct validity are Allshouse (2003) and Evans (2003). Allshouse's (2003) study concluded the ICIS instrument could accurately predict whether a teacher had high or low *knowledge of content*. Of the 41 teachers used in this study, the instrument correctly predicted a teacher's level of content knowledge, 100 percent of the time. Evans' (2003) study examined the ICIS sub-section of *works well with others.* This study concluded the ICIS was able to correctly predict a teachers' ability to work with others 80 percent of the time. Patrick Cowan (1999) is cited in the training manual due to the concurrent validity of his study. Cowan's (1999) study examined whether or not the ICIS could correctly identify effective versus ineffective teachers. The results showed an 86 percent rate of correctly identifying effective teachers and an 83 percent rate of correctly identifying less effective teachers as defined by principal ratings. Gary Stevenson (2007) conducted a recent study on whether teacher age or teaching experience can have an influence on the score received on the ICIS. Stevenson (2007) concluded that although age and experience may impact a candidate's score on the ICIS, it is probably not significant enough to warrant an adjustment to the candidate's interview score. Michael

Weishaar (2007) conducted a study on teacher effectiveness based on how their students performed on standardized tests compared to how the teacher was rated on the ICIS. Weishaar (2007) concluded that the ICIS interview could not effectively differentiate between teachers whose students performed at different levels on a given standardized test.

Finally, Michael Reik's (2007) study examined the relationship between a teacher's score on the ICIS and that same teacher's 3 year mean Missouri Assessment Program (MAP) test. The MAP test consists of a compilation of multiple-choice questions, constructed response, and performance events. Reik's research determined that there was a moderate to strong positive linear relationship between a teacher's score on the ICIS within the sub-sections of *Working with Others*, *Knowledge of Teacher*, and *Knowledge of Students* and the 3 year mean score on that same teacher's MAP. However, the sub-section of *Knowledge of Content*, and 3 year mean MAP proficiency percentages had a weak positive linear relationship (Reik, 2007).

A study on the predictive validity of the ICIS would provide further insights into its ability to "predict" teacher effectiveness. Presumably due to feasibility implications, no studies on the ICIS have been conducted exploring its predictive validity (rather than concurrent or criterion validity). To conduct a study on the predictive validity of the ICIS, data would have to be collected at one time and the outcome variable would be collected years later.

Principal Evaluations and Teacher Effectiveness

A secondary analysis included in this study was to examine the relationship between an administrator's rating of a teacher and that same teacher's effectiveness as defined by student achievement on the Kansas State Reading and Math Assessments.

A collection of studies in the education literature report relatively small correlations between principal evaluations and student achievement, although these studies are typically based on small, non-representative samples, often do not account properly for measurement error, and rely on objective measures of teacher performance that may be biased (Medley and Coker, 1987; Peterson, 1987; Peterson, 2000). According to a Harvard University study on the Measures of Effective Teaching (MET Project), many of the errors in teacher evaluations can be accounted for by the divergent perceptions between raters (Ho and Kane, 2013). It is not necessarily that some raters consistently rate teachers too high or too low (pg. 32). Ho and Kane (2013) go on to conclude that the only way a district can scrutinize the reliability of classroom observations and guarantee a fair and reliable system for teachers, would be to use multiple observers and set up a system to check and evaluate the feedback given to teachers by those different observers.

A number of recent findings have documented the substantial variation in teacher effectiveness within a school district and even within an individual school (Rockoff, 2004; Hanushek, Kain, O'Brien, and Rivkin, 2005; Aaronson, Barrow, and Sander, 2007). The impact this variation can potentially have on student success is significant. Rockoff (2004) estimates that the benefit of moving a student from an average teacher to one at the 85th percentile, is comparable to a 33 percent reduction in class size.

Additionally, Sanders and Rivers (1996) claim that the difference between having a series of very good teachers versus very bad teachers can be “enormous”.

Value-added assessment studies in Tennessee seem to indicate that the difference in achievement between students who attended classes taught by high-quality versus those taught by low-quality teachers for three consecutive years is sizeable: approximately 50 percentile points on standardized tests and that effective teachers are capable of inspiring significantly greater learning gains in their students when compared with their weaker colleagues (Sanders & Rivers, 1996). Some studies estimate how much value a teacher has contributed to student achievement, and factors in the gains the student was expected to make based on past performance (Crane, 2002). In Texas, economists have accumulated a body of work that further emphasizes the measurable influence that teachers have on student performance (Hanushek, Kain & Rivkin, 1998). What constitutes as teacher quality or effectiveness however is debatable.

Teacher quality can be extremely difficult to measure and therefore, most studies focus on measurable teacher outputs such as academic degrees, years of experience, and certifications. Some studies that have correlated teacher test scores on basic skills tests and college entrance exams with the scores of their students on standardized tests have found that high-scoring teachers are more likely to elicit significant gains in student achievement than their lower-scoring counterparts (Ferguson, 1998; Ferguson & Ladd, 1996; Strauss & Sawyer, 1986).

The literature shows that the employment interview continues to evolve as a tool for principals to determine the most effective teachers. Whether or not this evolution is

truly beneficial in and helping administrators make better decisions on which teachers are going to be most effective is still up for debate. This study looked at teacher effectiveness as measured by both student achievement and administrator ratings to determine whether or not the ICIS structured screening tool may be a beneficial tool for administrators to utilize when hiring teachers.

III. Methodology

The primary research question examined in this study, was if the ICIS displayed concurrent validity where the outcome variable was defined as student achievement on the Kansas State Reading and Math assessments. In addition, since extant research indicated significant correlations between the ICIS interview scores and administrator evaluations of the teacher (Cox, 2009; Dillon, 2006; Green, 2005; Hale, 2006; Smith, 2006) one would expect to find similar results in this study. Deviation from this finding would lead to a questioning of the design, data collection methods, excessive measurement error, or a host of other problems. The secondary research question for this study examined whether or not the ICIS had a correlation compared to an administrator's rating of a teacher on overall teaching effectiveness. As such, in a secondary analysis, administrators were asked to rate each of the 40 teachers in the study on a 1-10 scale regarding their overall teaching ability. Correlations were then calculated between ICIS interview scores and the administrator rating to see if the results replicated prior studies and also to help contextualize and understand the results from phase one of the analysis. To research this question, a correlation was calculated between an individual teacher's total weighted average score on the ICIS instrument to that same teacher's average three year score on the Kansas State Assessment in the areas of reading and math, after accounting for the effects of SES.

The Kansas State Assessments used in this study are designed to assess student understanding of various "indicators" of learning as defined by the Kansas State Department of Education (KSDE). The following is an example of a 4th grade reading indicator provided by KSDE: *Compares and contrasts information, (e.g., topics,*

characters' traits, themes, problem/solution, cause/effect relationships) in one or more appropriate-level text(s) and identifies compare/contrast signal words.

Subjects and Setting

The sample of this study consisted of 40 volunteer teachers from third-, fourth-, and fifth-grade. The teachers selected for this study came from 21 different elementary schools throughout the Olathe School District in Olathe, Kansas. Olathe School District is a large, relatively high SES school district, consisting of an enrollment of 28,228 students at the time of the study. The volunteer teachers selected must have served in their current position as a third-, fourth-, or fifth- grade teacher consistently at the same school for at least the past three years so the scores could be attached specifically to one teacher. During this time, students were selected and assigned to teachers in an effort to create a heterogeneous class for each teacher. Forms designed by Olathe Schools are provided to help homogenize the classes. Considerations taken into account include, but are not limited to: Title I services (math and reading support), behavioral needs, Individual Education Plans (IEP), and 504 Plans (accommodations for students with medical needs). Administrators and counselors use the information to determine classroom placement. A difference between buildings with regard to socio-economic status does, however, persist and therefore some teachers in the sample had higher or lower socio-economic status than the district average.

Each teacher was interviewed using the ICIS program and scored using a rubric based on their responses in the areas of Working with Others, Knowledge of Content, Knowledge of Teaching, and Knowledge of Students. The training module of the ICIS

requires the administrator to first become familiar with the four sub-sections of the ICIS (Working with Others, Knowledge of Content, Knowledge of Teaching, Knowledge of Students) by watching video clips of teacher responses to various interview questions posed. The training administrator is then asked to rate the response on a 1-3 scale based on what they believe the quality of that teacher's response to be. After the training administrator feels as though they are proficient in recognizing and appropriately scoring the quality of responses, they are required to complete a summative evaluation where several video clips of teacher responses from all four sub-sections are viewed and subsequently scored. The training administrator must indicate the same score on each teacher's video clip response as the ICIS program intended at least 90 percent of the time in order for that individual to be considered adequately trained.

To ensure there was no bias, the interviews for this study were conducted by a certified administrator not currently employed by Olathe School District and conducted prior to the collection of state assessment data.

The Interactive Computer Interview System

This study utilized the short version of the American Association Of School Personnel Administrator's (AASPA) Interactive Computer Interview System (ICIS) as its screening instrument. All four sub-sections of the ICIS were utilized including: *Working with Others, Knowledge of Instruction, Knowledge of Students*, and *Knowledge of Teaching*. The *Knowledge of Teaching* section consists of five subcategories: Delivery of Instruction, Planning, Interactions with Students, Assessment, and Climate

Development. See Appendix page 85 for a detailed description of all four components of the ICIS.

As stated previously in the literature review, the ICIS is a computer interview system that was created utilizing two primary documents: Teacher of the Future and Praxis III: Classroom Performance Assessments. Questions used for this interviewing instrument had to measure constructs represented in both documents. Additionally to assist with reliability, scoring rubrics for each question were designed. Finally, a computer program was written to assist in the presenting of the questions and the manipulating of the data once the questions had been scored.

The ICIS system utilizes a personal or laptop computer to allow the interviewer to focus on evaluating the candidate's responses in a face-to-face interview while the computer tracks response patterns, suggests potential questions based on the candidates responses, and ultimately creates a detailed summary report to capture various aspects of the interview (Ebmeier, 2003). The ICIS uses a branching interview technique in that once a specified level of standard deviation has been achieved in one of the four categories, the program no longer draws from that particular category. Reliability of each scale exceeds 0.90 and the validity of the instruments ranges from 0.25 to 0.80 given various dependent measures (Ebmeier, 2010).

Analysis

Once the teacher interview data had been collected using the ICIS program, performance data on Kansas State Assessments from that teacher's classroom was collected. Total percent correct scores in the area of reading and math for each class of

the participating teachers was provided by the Olathe School District's assessment department. Three-year mean scores for each teacher were then calculated for both reading and math, along with standard deviations to facilitate the calculation of a Z Score which removed the effect of test difficulty and length and made all the tests metrically equal.

Once Z Scores were calculated to normalize the data provided by the assessment department, the district's technology department provided socio-economic status information as defined by the percentage of students receiving free/reduced lunch in each teacher's class, for each of the three years this study examined. A residual score was then calculated in both reading and math with each teacher's assessment Z Score and free/reduced three-year average. This residual score calculated the adjusted test scores for each teacher that participated in this study, while accounting for the effect of socio-economic status (Table A4). These scores were then compared to the candidate's total weighted average on the ICIS to determine possible relationships using a correlation analysis.

A secondary analysis was also conducted since data had already been collected from administrator's regarding the estimated teaching effectiveness of each of the 40 teachers as part of the mini-validation study. Those scores were compared to that teacher's ICIS total weighted average, residual reading, and residual math score to determine possible relationships.

IV. Results

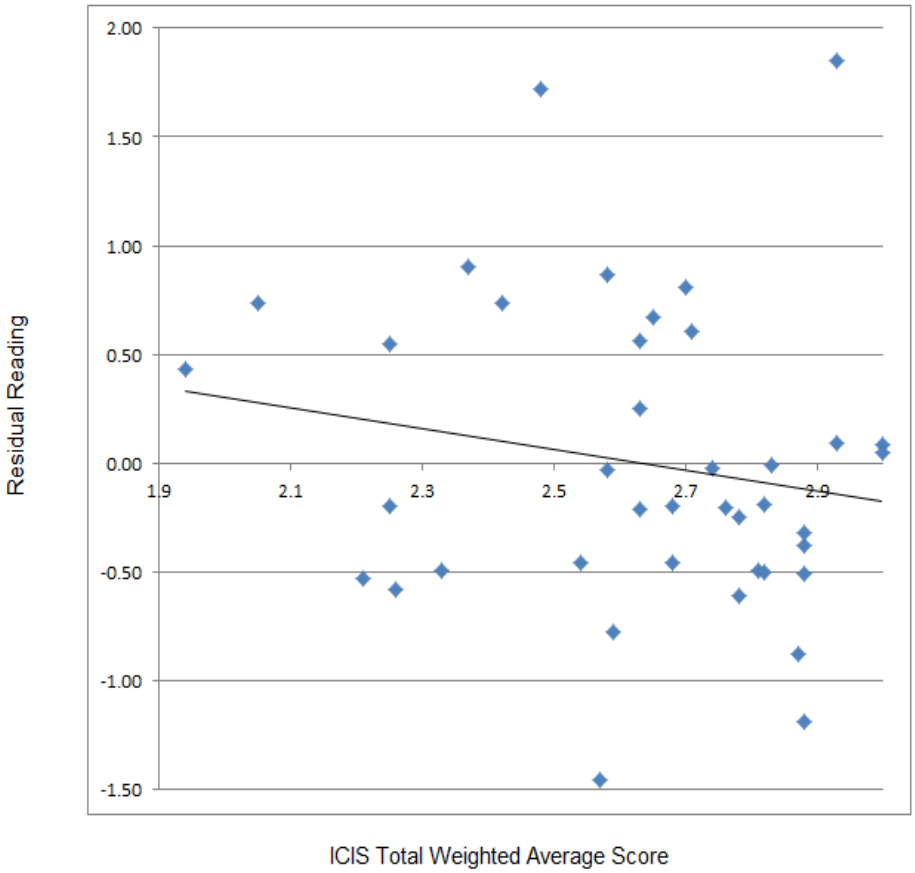
The primary purpose of this study was to examine the correlation between teacher interview scores on the Interactive Computer Interview System (ICIS) and that same teacher's student achievement based on the results of the Kansas State Assessment. Tables A4-A8, found in the Appendix, provide descriptions on the makeup of each individual teacher that participated in this study. Table A4 provides a description of the subjects including their gender, grade taught, and free/reduced lunch percentage. Table A5 provides raw data consisting of each participant's ICIS sub-section scores, total weighted average, residual math, and residual reading scores. Table A6 contains raw data of each teacher's total weighted average on the ICIS and their corresponding administrator rating. Table A7 consists of each teacher's total weighted average on the ICIS, administrator rating, residual reading, and residual math scores. Table A8 provides each teacher's mean score in reading and math on the Kansas State Assessments, along with their Z Score, Residual Reading and Residual Math.

Correlations of Primary Analyses

Graph 4.1 addresses the primary research question and displays a scatterplot illustrating the correlation between a teacher's total weighted average on the ICIS to that same teacher's residual reading score. Table 4.1 shows the descriptive statistics for this analysis and Table 4.2 shows the Pearson correlation. A very slight, negative correlation is found here, although the correlation is so small, it should not be inferred that as a teacher's total weighted average on the ICIS goes up, a teacher's student achievement in the area of reading goes down. There appears to be virtually no pattern

in this particular correlation. This data suggests there is no relationship between a teachers score on the ICIS and that teacher's student achievement on the Kansas State Reading Assessment.

Graph 4.1 Scatterplot of Relationship Between ICIS Total Weighted Average Score and Residual Reading



Correlaltion	-0.1782

Table 4.1: Descriptive Statistics for ICIS Total Weighted Average Score and Residual Reading

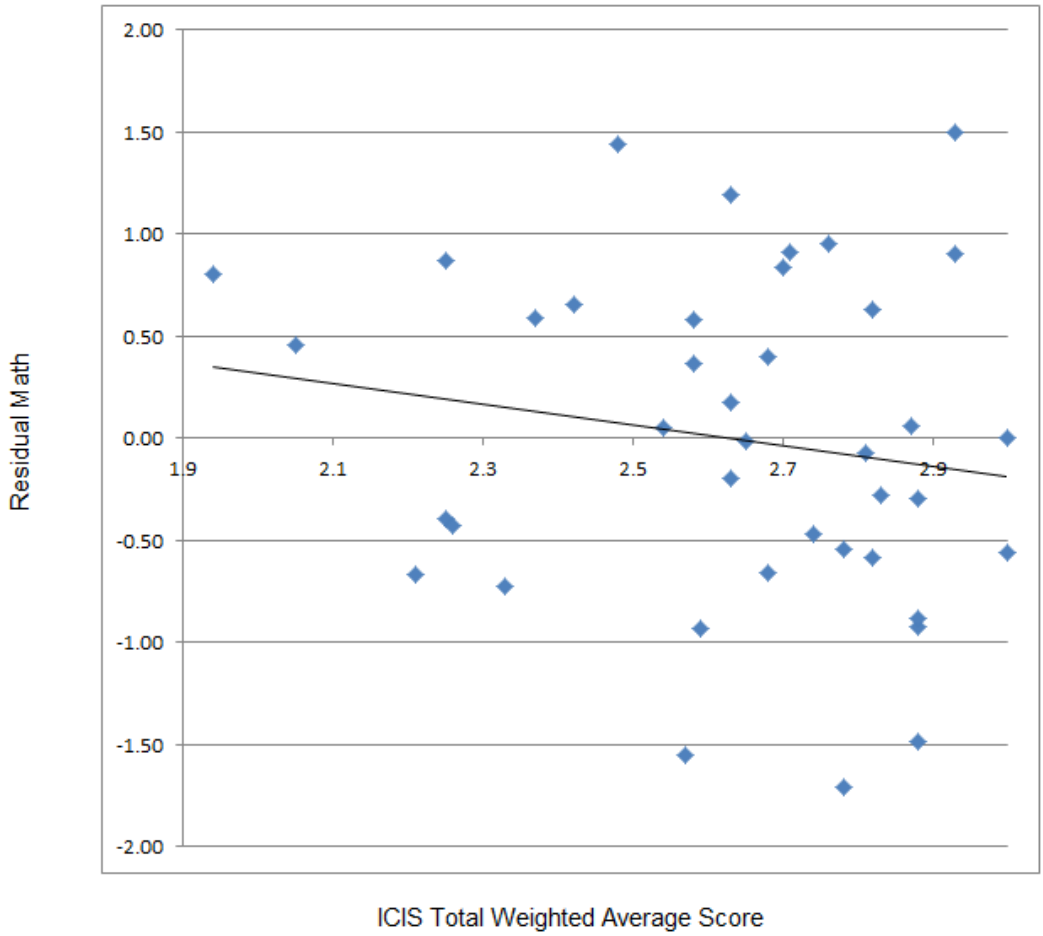
	Mean	**Standard Deviation**	**N**
ICIS Total Weighted Average Score	2.633	.23084	40
Residual Reading	-.0007	.70184	40

Table 4.2: Correlation for ICIS Total Weighted Average Score and Residual Reading

	ICIS Total Weighted Average Score to Residual Reading
Pearson Correlation	-.178
Sig. (2-tailed)	.276
N	40

Graph 4.2 also addressed the primary research question and displays a scatterplot illustrating the correlation between a teacher's total weighted average on the ICIS to that same teacher's residual math score. Table 4.3 is the descriptive statistics for this analysis and 4.4 is the Pearson Correlation. Similar to the correlation run between a teacher's total weighted average on the ICIS and that same teacher's reading score, there is no apparent pattern between a teacher's total weighted average on the ICIS and that same teacher's math scores. This data suggests there is no relationship between a teachers score on the ICIS and that teacher's student achievement on the Kansas State Reading Assessment.

Graph 4.2 Scatterplot of Relationship Between ICIS Total Weighted Average Score and Residual Math



Correlaltion	-0.1633

Table 4.3: Descriptive Statistics for ICIS Total Weighted Average Score and Residual Math

	Mean	**Standard Deviation**	**N**
ICIS Total Weighted Average Score	2.6330	.26084	40
Residual Math	.0002	.80961	40

Table 4.4: Correlation for ICIS Total Weighted Average Score and Residual Math

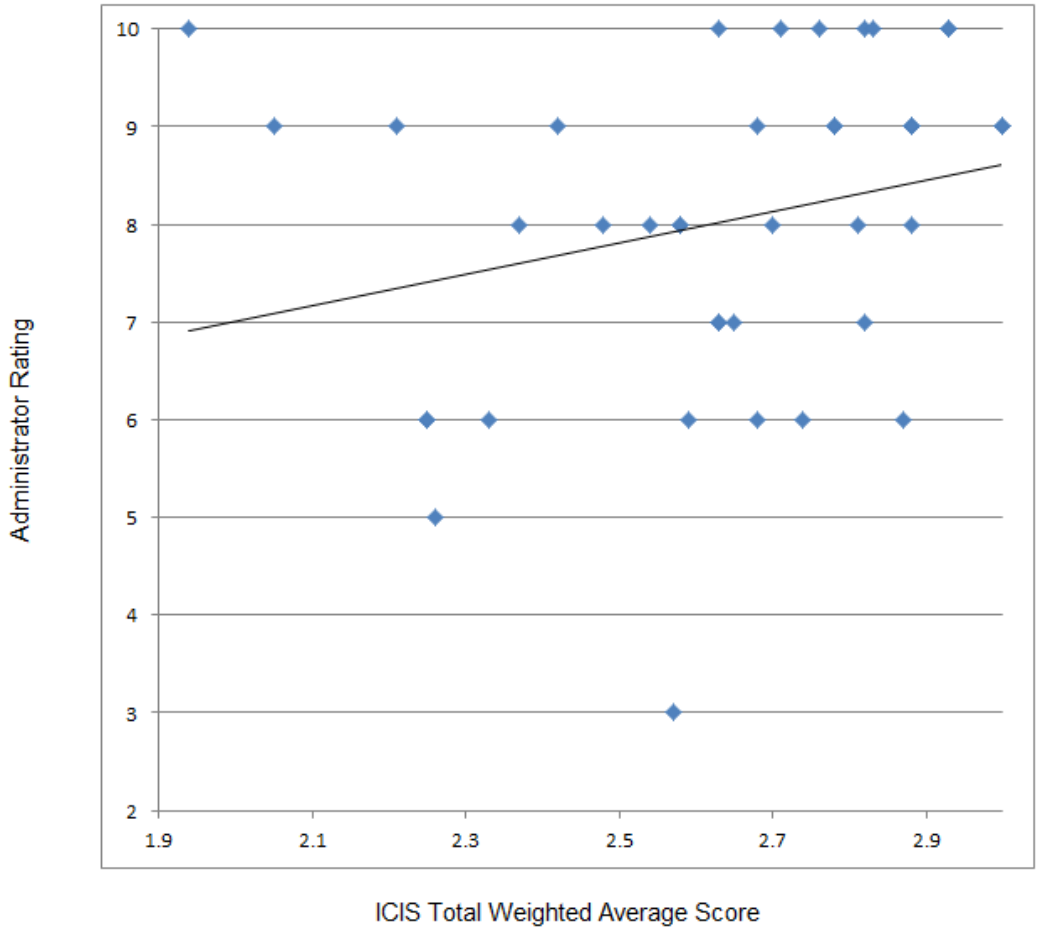
	ICIS Total Weighted Average Score to Residual Math
Pearson Correlation	-.163
Sig. (2-tailed)	.315
N	40

Correlations of Secondary Analyses

The secondary analysis for this study, was to examine any correlations that might exist between an administrator's rating of overall teaching ability to the total weighted average on the ICIS. Additionally, an examination of the relationship between an administrator's rating of a teacher and that same teacher's residual reading and math score was performed.

Graph 4.3 displays a scatterplot illustrating the correlation between a teacher's total weighted average on the ICIS to that same teacher's administrator rating on a scale of 1-10. Table 4.5 is the descriptive statistics for this analysis and 4.6 is the Pearson correlation. There does appear to be a slight correlation between these two variables with a correlation of 0.2498. The p-value of .120 suggests there may some degree of statistical significance between these two variables.

Graph 4.3 Scatterplot of Relationship Between Administrator Rating and ICIS Total Weighted Average Score



Correlaltion	0.2498

Table 4.5: Descriptive Statistics for ICIS Total Weighted Average Score and Administrator Rating

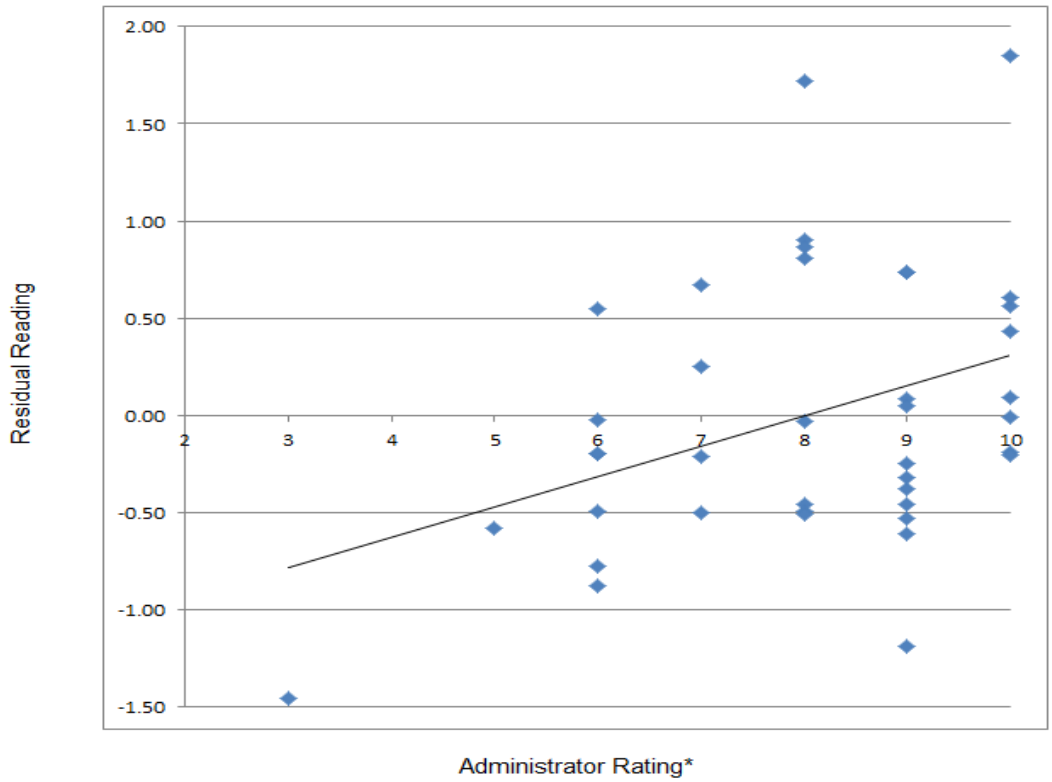
	Mean	**Standard Deviation**	**N**
ICIS Total Weighted Average Score	2.6330	.26084	40
Administrator Rating	8.03	1.672	40

Table 4.6: Correlation for ICIS Total Weighted Average Score and Administrator Rating

	ICIS Total Weighted Average Score to Administrator Rating
Pearson Correlation	.2498
Sig. (2-tailed)	.120
N	40

Graph 4.4 displays a scatterplot illustrating the correlation between a teacher's administrator rating and that same teacher's residual reading score. Table 4.7 is the descriptive statistics for this analysis and 4.8 is the Pearson correlation. This relationship had a statistically significant relationship at the .05 level with a correlation of 0.3718. This data suggests that an administrator may be a good evaluator of teacher effectiveness as defined by student achievement on the Kansas State Reading Assessment.

Graph 4.4 Scatterplot of Relationship Between Administrator Rating and Residual Reading



Correlaltion	0.3718

*Administrator Criterion:
10- Exemplary
7- Above Average
5- Average
3- Below Average
1- Poor

Table 4.7: Descriptive Statistics for Administrator Rating and Residual Reading

	Mean	**Standard Deviation**	**N**
Administrator Rating	8.03	1.672	40
Residual Reading	-.0007	.70184	40

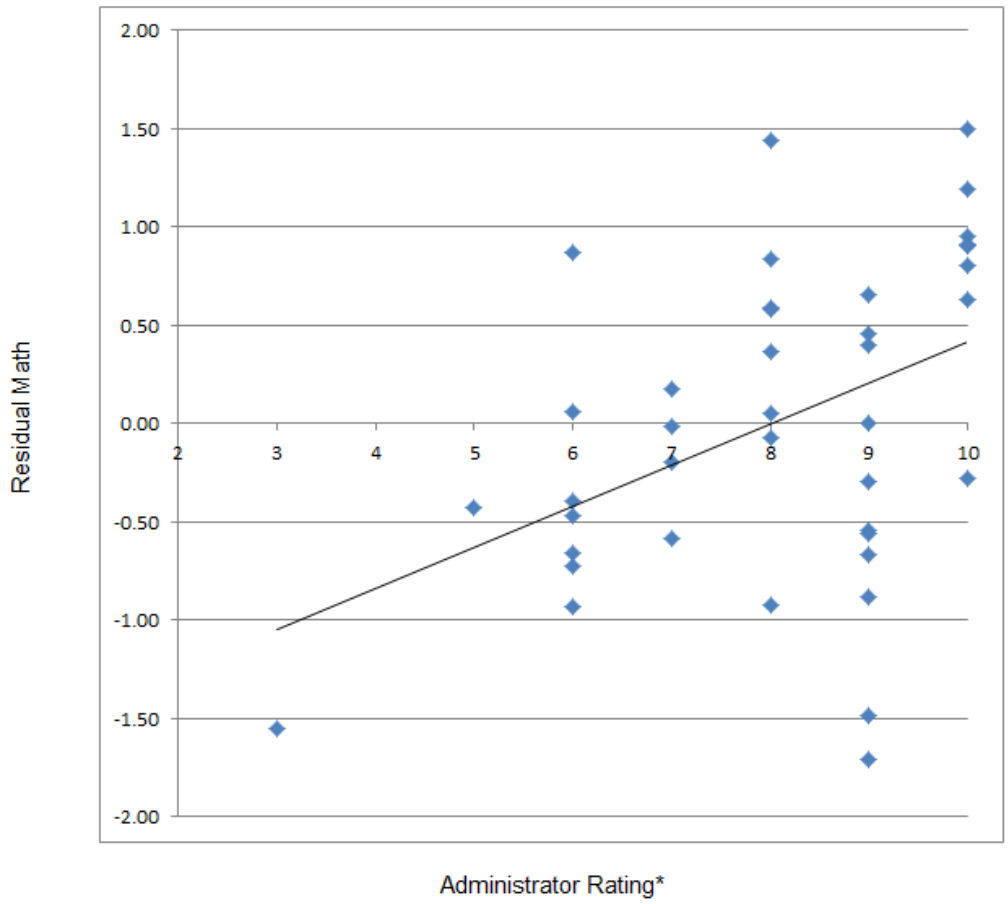
Table 4.8: Correlation for Administrator Rating and Residual Reading

	Administrator Rating to Residual Reading
Pearson Correlation	.3718*
Sig. (2-tailed)	.018
N	40

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

Graph 4.5 displays a scatterplot illustrating the correlation between a teacher's administrator rating and that same teacher's residual math score. Table 4.9 is the descriptive statistics for this analysis and 4.10 the Pearson correlation. Similar to the relationship between a teacher's administrative rating and residual reading score, a correlation of 0.4309 was found between a teacher's administrative rating and residual reading score. This correlation is statistically significant at the .01 level. This data suggests that an administrator may also be a good evaluator of teacher effectiveness as defined by student achievement on the Kansas State Math Assessment.

Graph 4.5 Scatterplot of Relationship Between Administrator Rating and Residual Math Adjustment



Correlaltion	0.4309

*Administrator Criterion:
10- Exemplary
7- Above Average
5- Average
3- Below Average
1- Poor

Table 4.9: Descriptive Statistics for Administrator Rating and Residual Math

	Mean	**Standard Deviation**	**N**
Administrator Rating	8.03	1.672	40
Residual Math	.0002	.80961	40

Table 4.10: Correlation for Administrator Rating and Residual Math

	ICIS Total Weighted Average Score to Residual Math
Pearson Correlation	.431**
Sig. (2-tailed)	.005
N	40

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

In summary, there appears to be a slight correlation (0.2498) between a teacher's total weighted average on the ICIS and that same teacher's administrator rating. A slightly higher correlation of 0.3718 was found to exist between an administrator's rating of a teacher and that same teacher's residual reading score. This relationship is statistically significant at the .05 level. An even higher correlation of 0.4309 was found to exist between an administrator's teacher rating and that same teacher's residual math score. This relationship is statistically significant at the .01 level.

Table A9 in the Appendix displays detailed correlation information for the sample, including a correlation for each sub-section of the ICIS and how it correlates to the residual reading and math. It should be noted that none of the ICIS sub-sections have a statistically significant correlation to either residual reading or residual math.

V. Discussion

Primary Analysis

While a number of studies have been conducted on the Interactive Computer Interview System (Allshouse, 2003; Evans, 2003; Cowen, 1999; Weishaar, 2007; Cox, 2006; Dillon, 2006; Dugan, 2007; Green, 2005; Longnecker, 2006; Smith, 2006), few have examined the concurrent validity as it pertains to student success on a criterion-referenced test (Reik, 2007).

This study found no statistically significant correlation between a teacher's total weighted average on the ICIS and that same teacher's average residual reading score and average residual math score on the Kansas State Assessments from 2010, 2011, and 2012. An examination of the scatter plots illustrating this correlation in Chapter 4 (Graphs 4.1, 4.2), shows no apparent pattern, although both Pearson correlations are very similar, only having a difference of 0.149 between the two. With both reading (-0.1782) and math (-0.1633) having such a minimal negative correlation, it should not be inferred that as a candidate's total weighted average on the ICIS gets higher, a student's achievement gets lower. Rather, it may be inferred that there is simply no relationship between these variables. It should also be noted that none of the four sub-sections of the ICIS had a statistically significant correlation as it related to student achievement on the Kansas State Assessment.

Teachers #13 and #5 had the lowest scores on the ICIS of 1.94 and 2.05 respectively. If these two scores are removed from the analysis between total weighted average and reading, the correlation becomes a virtually non-existent -0.0875.

Likewise, if these scores are removed from the analysis between ICIS total weighted average and math achievement, this correlation also becomes a virtually non-existent -0.0721. This data suggests that concurrent validity does not exist between a teacher's performance on the ICIS screening tool and a criterion-referenced test such as the Kansas State Assessments.

These findings differ from the similar study conducted by Reik (2007) where 40 third- and fourth- grade teachers were interviewed using the ICIS screening tool and those scores were correlated to student achievement on the Missouri Assessment Program (MAP). Reik found that there was a statistically significant correlation between these two variables (0.454). One possible explanation for the difference between the results of these two studies could be the differences between the outcome variable of the MAP Assessment and the Kansas State Assessments. The MAP Assessment utilizes a combination of selected-response, multiple choice, constructed-response, and open-ended essay questions (DESE, 2003), whereas the Kansas State Assessment simply uses multiple-choice questions.

The current findings would however be consistent with the study conducted by Weishaar (2007), although Weishaar used standardized test data as the dependent variable as opposed to criterion-referenced data. In this study, Weishaar concluded that the ICIS interview did not effectively differentiate between teachers whose students performed at different levels on a given standardized test. Other studies using student achievement as the outcome variable with differing employment selecting instruments have generally also found no relationship between the selection instrument and subsequent student academic achievement (Hall, 2010; Regan, 2011; Kirchner, 2008).

This lack of apparent relationship is likely attributable to several conditions in the experimental design, sample section, or quality of the data as discussed later.

Secondary Analysis

The secondary analysis of this study examined the correlation between a teacher's total weighted average on the ICIS and that same teacher's administrative rating. This relationship showed a slightly positive, yet not statistically significant correlation of 0.2498. It may be interesting to note however, that two outliers may have considerably impacted the results of this particular analysis. In this section, the teacher assigned #13 scored a 1.94 on the ICIS and a ten on the administrator rating. Additionally, teacher #17 scored a 2.57 on the ICIS received a three on the administrator rating scale. If these scores were removed, the correlation would jump to a statistically significant 0.4001. A larger sample size would be necessary to determine which correlation is more indicative of the predictive validity of the ICIS as it relates to administrative rating.

The p-value for this correlation is another important factor for consideration due to the relatively small sample size of 40. Using only the traditionally considered values of statistical significance (.05 and .01) may not give an accurate picture of the true significance of these correlations. In this instance, the p-value for the correlation between ICIS score and administrator rating was 0.120. This p-value could be perceived as statistically significant (at the .12 level), indicating there is in fact little chance the results of this section of the study merely occurred by chance.

The final examination of this study was to explore the possible relationship between a teacher's administrator rating and that same teacher's residual reading and math scores. There was a statistically significant correlation at the .05 level between an administrator's rating and residual reading of 0.3718 and an even larger correlation at the .01 level between an administrator's rating and residual math of 0.4309.

Similar to the earlier analysis examined between a teacher's total weighted average on the ICIS to that same teacher's student achievement data, the relationship between a teacher's administrative rating compared to student achievement may also have been impacted by a few outliers. For example, teacher # 28 scored a -1.19 on their residual reading score, but was given a rating of nine by their administrator. If this one outlier is removed, the correlation between a teacher's administrative rating and that same teacher's residual reading score jumps from 0.3718 to 0.4156. Likewise, the teachers assigned #28 and #23 had a significant impact on the correlation between a teacher's administrative rating and that same teacher's student achievement in the area of math. Teacher #28 scored a -1.48 on their residual math, but a nine from their administrator, while teacher #23 scored a -1.71 on their residual math, but a nine from their administrator. If these two outliers are removed, the correlation between administrative rating and residual math jumps from 0.4309 to 0.5599. This data may imply that an administrator's knowledge of their own teacher's overall teaching ability is a better predictor of student success on a criterion-referenced test such as the Kansas State Assessment than a screening tool such as the ICIS. In order to examine this possibility closer, a larger sample size would be necessary. It should be noted that the request made to each building principal was to rate their teachers that participated in

the study by “overall teaching effectiveness.” Results may have been different had principals been asked to rate their teachers based on how they felt they prepared students for Kansas State Assessments.

The results of this study support one conducted by Jacob and Lefgren (2008) who examined whether or not a principal had the ability to identify a teacher’s ability to increase reading and math achievement. The results of Jacob and Lefgren’s study indicated that principals are in fact generally effective at identifying the very best and very worst teachers. However, principals were not nearly as successful at identifying teachers that were in the middle of the achievement distribution. This study also went on to conclude that principal ratings were a significant predictor of future student achievement (Jacob & Lefgren, 2008). This is commensurate to the results of this student where principal ratings were found to be a relatively reliable indicator of student achievement in both reading and math.

Another consideration would be the use of average residual scores from the 2010, 2011, and 2012 state assessments. Had this study looked at all residual reading and math scores from each of the three years in which data was collected, the number of data points examined would have jumped to 120 (as opposed N=40). Examining the data in this way would have only slightly altered the correlations. Table 5.1 shows what each of the correlations were using average residual reading and math scores, as compared to using each individual residual reading and math score (N=120). This slight adjustment in correlations implies that a larger sample size may not have impacted the results of the primary analysis of the study.

Table 5.1 Comparison of Average Scores to Individual Scores

Category	Correlation Using 40 Averaged Scores	Correlation Using 120 Individual Scores
ICIS Total Weighted Average to Residual Reading	-0.1782	-0.1341
ICIS Total Weighted Average to Residual Math	-0.1633	-0.1180
Administrator Rating to Residual Reading	0.3718*	0.2687
Administrator Rating to Residual Math	0.4309**	0.3411*

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

Another consideration resulted from examination of data provided by Olathe School District and the Kansas State Department of Education. The data provided showed a much greater level of student achievement by teachers in Olathe as compared to teachers in Unified School District's throughout the state of Kansas. For example, the mean raw score for third graders in the area of reading throughout the State of Kansas from 2010, 2011, and 2012 showed the total percent of questions answered correct was 79.86 percent. Comparatively, according to Olathe School District's assessment department, the mean raw score for total questions answered correctly by third grade students over the same time period throughout Olathe was 91.96 percent. This means Olathe students scored better than the state mean by over 12 percentage points.

Limitations

Variability in each teacher's achievement test scores over time may in fact have been a hindrance for the primary analysis of this study. Tables A10-A15 in the Appendix show the variability between residual reading scores and residual math scores based on quartiles. Table A10 consists of each grade level's minimum,

maximum, and quartile residual reading scores. Table A11 provides the same information as A6, but for math residual scores. Table A12 outlines the residual scores, organized by grade and year. Table A13 provides the quartile ranking (1-4) for each teacher, organized by grade taught and year. Tables A14 and A15 display the variance between quartile rankings for each of the 40 teachers utilized in this study for reading and math respectively.

After teachers' scores were normed and adjusted to remove the effects of socio-economic status, teachers were ranked by quartile for each subject and year. Stability of scores was tracked by quartile rankings from year to year, showing the percentage of teachers with stable scores from year to year and those who changed quartiles from one year to the next. Table 5.2 below shows the breakdown of tracked teacher reading movement from one year to the next. The findings indicated in reading that teacher movement by at least one quartile to another occurred an average 68.75 percent of the time (from Year 1 to Year 2 and Year 2 to Year 3). Conversely, teachers stayed in the same quartile in reading from one year to the next an average of 31.25 percent of the time.

In the area of math, the findings indicated teacher movement by at least one quartile or more occurred with an average of 62.5 percent of the teachers (Table 5.2). An average of 37.5 percent of teachers stayed in the same quartile from one year to the next. The least often occurring teacher score movement in both reading and math was movement two or more quartiles away (i.e. first quartile to third quartile), Movement by more two or three quartiles would show the most variability of teacher stability.

Table 5.2: Teacher Quartile Movement Over Time

Teacher Movement Over Time- Reading	**Yr 1 to Yr 2**	**Yr 2 to Yr 3**	**Mean**
Same Quartile	42.5%	20%	31.25%
Movement to Quartile Above/Below	47.5%	52.5%	50%
Movement Two Quartiles Away	10%	27.5%	18.75%
Teacher Movement Over Time- Math	**Yr 1 to Yr 2**	**Yr 2 to Yr 3**	**Mean**
Same Quartile	40%	35%	37.5%
Movement to Quartile Above/Below	50%	45%	47.5%
Movement Two Quartiles Away	10%	20%	15%

Teacher movement in reading between quartiles showed similar variability from year to year in both reading and math. In both subjects, the highest mean percentage movement was in the category of moving up or down by one quartile (50 percent in reading; 47.5 percent in math). The next highest category in both reading and math was no movement between quartiles (31.25 percent in reading; 37.5 percent in math). The lowest percentage of movement was in the category of movement by two or more quartiles (18.75 percent in reading; 15 percent in math). The instability of these scores may suggest that the Kansas State Assessments are not a reliable indicator of teacher effectiveness. While these assessments are widely considered rigorous and useful in determining a student's ability to "know and do" agreed upon standards, it is only one summative assessment for reading and one for math, given in two weeks, of one school year. Using one such assessment to determine a teacher's overall performance is unfair. However, the current study may provide a starting point for a more complete approach to establishing teacher effectiveness and in turn, evaluate instrument validity. This is especially true as more school districts consider tying student performance into teacher evaluation.

Other recent studies have also indicated that instability in student learning gains is often present with top performing and bottom performing teachers, although these studies used quintiles as opposed to quartiles (Koedel & Betts, 2007; McCaffrey, Lockwood, Koretz, Hamilton, 2003). In total, their results showed an average of 12 percent of the lowest and highest performing teachers moving up from bottom quintiles or down from top quintiles. McCaffrey et al. (2003) found instability in an analysis of five urban school districts in various parts of the country. These research findings indicated that, of the teachers ranked in the bottom quintile of teacher effectiveness in one year, only 25-35 percent were ranked in the same quintile the following year. Of teachers in the top quintile one year, only 20-30 percent were ranked in the same quintile a year later, while about the same percentage of teachers fell to the bottom quintile of performance.

An additional limitation that may have contributed to the results found in the primary analysis of this study could be the quality of the teachers utilized. The Olathe School District's third-, fourth- and fifth- grade teachers that participated in this study averaged a raw score on the ICIS of 2.63 which falls into the 78th percentile of scores on the ICIS. Although these 40 teachers volunteered and were not "recommended teachers" by administrators, the average administrative rating for all 40 teachers was 8.025 (on a scale of 1-10). These high administrative ratings would suggest that the teachers utilized in this study may not be your "typical" teachers and therefore would not be the most representative sample of teachers throughout the state of Kansas.

Another possible limitation exists in the practical application of concurrent validity. According to Cherry (2010), "concurrent validity" occurs when the criterion

measures are obtained at the same time as the test scores. This indicates the extent to which the test scores accurately estimate an individual's current state with regards to the criterion. Perhaps a more useful approach would be to establish predictive validity (administer the interview at one date and use student achievement scores attained over a period of time following the interview). Cherry (2010) describes "predictive validity" as when the criterion measures are obtained at a time after the test. Such a study presents complications due to the time commitment associated with multi-year studies (Ebmeier, 2006). If concurrent validity is established in a structured interview, it may be reasonable for an interviewer to cautiously assume some predictive qualities are present given the limitations. As previously discussed in the literature review, a structured interview, in general, can help in predicting job performance (Huffcutt, Conway, Roth, & Stone, 2001).

The number of years of teaching experience and how that number relates to teacher effectiveness may also have had an impact on this study. All of the teachers interviewed in this study had at least three years of teaching experience. Stakeholders would like to believe that a teacher continues to improve each year until retirement, however recent research seems to suggest that teachers may actually plateau after about five years of teaching experience (Clotfelter, Ladd, & Vigdor, 2006). Although many school systems tend to promote the concept that teaching experience is paramount to student success, first-year teachers are often assigned the same responsibilities as a twenty-year teaching veteran. Instructors may improve with experience, but it's possible that other variables that affect a teacher's attitude and

motivation throughout his or her career, and experience can only be considered as a positive influence to a certain point.

Interviewer bias may also have been a limitation with this study. Even though the interviewer was a retired administrator, not associated with the Olathe School District, adequately trained on using the ICIS screening tool, some coding drift is only natural. This drifting may be particularly true given that all 40 interviews were completed over a period of about two weeks.

A final limitation would be to consider the primary function of the ICIS. The ICIS is not designed to predict which teachers will have positive test scores, but rather is intended to be a resource for experts such as administrators and Human Resource Directors in determining which candidates will have the best chance to be successful in the four areas evaluated with the ICIS: Working with Others, Knowledge of Content, Knowledge of Teaching, and Knowledge of Students (Ebmeier, 2001). This could account for the reason this study showed no relationship between a teacher's score on the ICIS as compared to student achievement, but did show a positive relationship between a teacher's score on the ICIS and an administrator's rating.

Conclusions

The primary analysis of this study was to determine whether the ICIS screening tool was able to determine teacher effectiveness as defined by student achievement on the Kansas State Assessments in both reading and math. As detailed in the "Results" section, there appeared to be no correlation between these two variables. However, a number of variables outlined in the "Discussion" and "Limitations" section of this study

such as a teacher's quartile inconsistency and comparison data between the state mean and Olathe School District mean, indicate that student achievement on these state assessments may not have been a valid dependent variable and therefore, may not have been a good indicator of teacher effectiveness. This is perhaps one of the primary reasons so many other studies examining the effectiveness of commercial screening tools focus on administrator ratings and not student achievement data (Sawyer, 2005; Delli & Young, 2002; Green, 2005; Smith, 2006).

The secondary analysis, which examined the correlation between a teacher's ICIS score to an administrative rating, did result in a potentially significant relationship worth exploring more. With regard to practical application, a strong correlation between an administrator's rating and a commercial screening tool such as the ICIS could potentially be more appealing to a school district than that screening tool's ability to accurately predict which teachers will high test scores.

The final piece of the secondary analysis was to examine the correlation between an administrator's rating of a teacher to that same teacher's student achievement on the Kansas State Assessments in both reading and math. These correlations turned out to be the highest of any of the correlations run in this study (0.3718 in reading, 0.4309 in math). This data may suggest that an administrator knows their teachers better than a commercial screening tool. This conclusion would certainly have practical implications due to the fact that administrators typically do not have the opportunity to observe a potential staff member for a period of time before determining whether or not they would

be a good fit with their staff. Perhaps a slightly more feasible consideration would be to build in a short performance-based component to a structured interview process.

Suggestions for Future Research and Improvements

The definition for teacher effectiveness in this study was defined by student performance on the Kansas State Reading and Math Assessment. These tests, taken each year in the Spring, have remained relatively consistent since their introduction back in 2003. The sample group of teachers in this study have taught in their current grade level and their current building for at least the past three years. This means every teacher that was part of this sample has given virtually the same assessment for at least the past three years. In the future, it may be beneficial to examine data that could measure student growth from the beginning of the year to the end of the year. Many districts, including Olathe School District, have Beginning of Year and End of Year Assessments that are intended to measure growth from the beginning of the year to the end of the year. One could argue that student growth over the period of an academic year is more indicative of teacher effectiveness than a criterion-referenced summative assessment such as the Kansas State Assessments.

It may also be beneficial to examine a larger sample size, consisting of teachers from other schools districts, rather than limiting the study to teachers from one school district. While there is some diversity in Olathe, all 40 of the teachers that participated in this study were Caucasian and taught in an elementary school. A larger sample size made up of ethnically diverse teachers from a variety of school districts would be worth examining.

Another consideration for future studies might be to explore the predictive validity of the ICIS screening tool, as opposed to its concurrent validity. A number of studies have already explored the concurrent validity of the ICIS (Allshouse, 2003; Evans, 2003; Reik, 2007; Cowan, 1999). One possible way to examine the predictive validity of the ICIS would be to administer the ICIS screening tool to new teachers hired within a school district for several years and then compare student achievement data and administrator ratings for those same teachers several years in the future. This study would have logistical implications due to the length of time it would take to acquire this data. A screening tool that had research supporting its predictive validity would certainly be appealing to a school district as predicting the future success of an educator is one of the primary objectives of any school district's human resources department.

References

Allshouse, Tim (2003). Construct validity of the knowledge of content scale from the AASPA Interactive Computer Interview Instrument, Unpublished dissertation, University of Kansas, Lawrence, KS.

Antoline, C.M. (2000). A validation study of the Teacher Perceiver Interview (Doctoral dissertation , Texas Women's University, 2000).

Applegate, J. (1987). Teacher Candidate Selection: An Overview. *Journal of Teacher Education* 38, 2-6.

Arrestad, J. (1980). Analysis of teacher selection/retention and student achievement as related to Selection Research Incorporated's Teacher Perceiver Interview, Doctoral dissertation, University of Minnesota, Minneapolis, MN.

Arvey, R.D., Campion, J.E. (1982). "The employment interview: A summary and review of recent research." Personnel Psychology. v35, 281322.

Asch, S.E. (1952). Social Psychology. Englewood Cliffs, New Jersey: Prentice Hall, 1952.

Baker, H.D., & Morris, S.S. (1990). The employment interview. Guaranteed improvement in reliability. Public Personnel Management,19, 8590.

Baskin, M.K., Ross, M.R., Smith, D.L. (1996). "Selecting successful teachers: The predictive validity of the urban teacher selection interview." The Teacher Educator, v32, 121.

Best, R. (2005). *An Exploration of Preservice Teachers And The Functions Of Star Teachers.* Doctoral dissertation, University of Dayton, Dayton, Ohio.

Blank, R.K., "Meeting NCLB Goals for Highly Qualified Teachers: Estimates by State from Survey Data," Council of Chief State School officer, 2003, October.

Bond, L., Smith, T.W., Baker, W. K., & Hattie, J.A. (2000). *The certification system of the National Board for Professional Teaching Standards: A construct and consequential validity study*. Greensboro, NC: Center for Educational Research and Evaluation.

Caldwell, T. (1993). Hiring excellent teachers: Current interviewing theories, techniques, and practices. Long Beach: California State University (M.A. Project).

Carlson, R., Thayer, P., Mayfield, E., & Peterson, D. (1971). Improvements in the selection interview. *Personnel Journal*, 50, 268-275.

Carney, M. & Johnson, E. (1995). Evaluation of Haberman Urban Teacher Selection Interview in the Minneapolis Public Schools. Minneapolis, MN: Minneapolis Public Schools Human Resources Department.

Castetter, W., & Young, I. (2000). *The human resource function in educational administration*. Columbus, OH: Prentice Hall.

Cavaluzzo, L. C. (2004). *Is National Board Certification an effective signal of teacher quality?* Alexandria, VA:The CNA Corporation. Retrieved from http://www.nbpts.org/policy_center/education_policy_prek-1/redefining_teacher_quality

Chalker, Donald M. (1981). *The Teacher Perceiver Interview as an instrument for predicting successful teaching behavior*. Dissertation Abstracts International, 42, 02A (UMI No. 8117049).

Cherry, Kendra (2010). *The everything psychology book: 2nd edition*. Avon, MA: Adams Media

Clement, M. C. (2009). Hiring highly qualified teachers begins with quality interviews. Kappan. 9(2), 22-24.

Clotfelter, Ladd, and Vigdor (2006). *Teacher-student matching and the assessment of teacher effectiveness*. National Bureau of Economic Research.

Coleman, J., Campbell, E., Hobson, C., McPartland, J., Mood, A., Weinfeld, F., & York, R. (1966). *Equality of educational opportunity*. Washington, DC: U.S. Government Printing Office.

Conway, Jako and Goodman DE (1995). A metanalysis of interrater and internal consistency reliability of selection interviews. *Journal of Applied Psychology*, 80, 565-579.

Cowan, P. (1999). *A comparison of the predictive power of competency-based and Personality based structured interviews in identifying successful teachers*. Doctoral dissertation, University of Kansas, Lawrence, KS.

Cox, D. (2006). *A study of the reliability of the aptitude/suitability for teaching in urban schools scale from the American Association of School Personnel Administrators (AASPA) Interactive Computer Interview System (ICIS)*. Doctoral dissertation, University of Kansas, Lawrence, KS.

Darling-Hammond, L. (2000). Teacher quality and student achievement: A review of state policy evidence. *Education Policy Analysis Archives, 8*(1).

Delli, A. D., (2001). *The predictive validity of the teacher perceiver interview on the teacher performance of classroom teachers*. Unpublished doctoral dissertation. The Ohio State University.

Delli, D. A., & Vera, E. M. (2003). Psychological and contextual influences on the teacher selection interview: A model for future research. Journal of Personnel Evaluation in Education, 17(2), 137-155.

Delli, D.A. & Young, I.P. (2002). "The validity of the teacher perceiver interview for predicting performance of classroom teachers". *Educational Administration* Quarterly, 38(5), 586-613.

Dillon, A. (2006). The development of an interview system to identify quality paraprofessionals. Doctoral dissertation. University of Kansas, Lawrence, KS.

Doll, D. (2009) *The Correlation Between the Haberman Star Teacher Selection Interview and the AASPA Interactive Computer Interview System – Urban*. Doctoral dissertation. University of Kansas, Lawrence, KS.

Dugan, E. (2007). *The reliability of building administrators' and special education administrators' ratings of special education teacher applicants using an Interactive Computerized Interview System*. Doctoral dissertation, University of Kansas, Lawrence, KS.

Ebmeier, H. & Ng, J. (2005). Development and field test of an employment selection instrument for teachers in urban school districts. Journal of Personnel in Education, v.18, n.3, September 2005, pp. 201218.

Ebmeier, H. (2006). Interactive Computer Interview System Technical Manual. American Association of School Personnel Administrators, V3.

Eder, R.W. & Ferris, G.R., (1989). The Employment Interview: Theory, Research, and Practice. Sage Publications: Newbury Park, CA.

Emley, K., & Ebmeier, H. (1997). The effect of employment interview format on principals' evaluations of teacher. *Journal of Personnel Evaluation in Education*, 11, 39-56.

Engle, R., & Erion, L. (1984). A comprehensive review of interviewing strategies for the 80's. *Clearinghouse*, 57, 304-308.

Evans, L. (2003). *Construct validity of the working with others scale from the AASPA Interactive Computer Interview Instrument*, Doctoral dissertation, University of Kansas, Lawrence, KS.

Ferguson, R. (1991). Paying for Public Education: New Evidence on How and Why Money Matters. *Harvard Journal of Legislation* 28: 465–98.

Ferguson, R.F. (1998). "Teachers' Perceptions and Expectations and the Black-White Test Score Gap," In Christopher Jencks and Meredith Phillips (Eds.), *The Black-White Test Score Gap* and *Can Schools Narrow the Black-White Test Score Gap?* pp. 273-374, Washington, D.C.: Brookings Institution Press.

Ferguson, R.F. and Ladd, H.F. (1996). "How and Why Money Matters: An Analysis of Alabama Schools," In Helen F. Ladd (Ed.), *Holding Schools Accountable: Performance Based Reform in Education*, Washington, D.C.: Brookings Institution Press, 1996.

Fisher, S. & Dickenson, T. (2005). *A study of the relationship between the National Board Certification status of teachers and students' achievement*. Technical report prepared for the South Carolina State Department of Education. Study conducted by the College of Education, University of South Carolina. Retrieved from www.educationconsumers.com/.../South%20Carolina%20NBPTS%20report.pdf

Gallup Organization (1997). The teacher perceiver: Overview, background, and research (brochure). Lincoln, NE: Author.

Gallup Organization. (2009). The TeacherInsight. Retrieved April 25, 2009, http://www.gallup.com/consulting/education/22093/TeacherInsight.aspx

Gillies, T. (1988). The relationship between selection variables and subsequent performance ratings for teachers in an Oregon school district. Doctoral dissertation, University of Oregon, Eugene, OR.

Goldstein, W. (1986). Recruiting superior teachers: The interview process. *Phi Delta Kappan,* Bloomington, IN: Phi Delta Kappa Educational Foundation.

Green, M.E. (2005). *The development and field testing of an interview instrument designed to identify quality secondary school counselors*. Doctoral dissertation. University of Kansas, Lawrence, KS.

Haberman, M. (1993). *Predicting the success of urban teachers (the Milwaukee trials). Action in Teacher Education*, 15 (3), 15.

Haberman, M. (1995). *Selecting 'star' teachers for children and youth in urban poverty. Phi Delta Kappan*, 76 (10), 777781.

Haefele, D.L. (1978). The teacher perceiver interview: How valid? Phi Delta Kappa, 59 (11), 683-684.

Hall, Jon (2010). *Fourth and fifth grade student achievement and possible relationship to three teacher selection tools*. Doctoral dissertation. University of Louisville, Louisville, KY.

Hanushek, E.A., Kain, J.F., and Rivkin, S.G., (1998). Teachers, schools, and academic achievement," (NBER Working Paper No. w6691), *National Bureau of Economic Research*.

Harvey, V., & Struzziero, J. (2000). Effective supervision in school psychology. Bethesda, MD: NASP.

Ho, A., Kane, T. (2013). *The reliability of classroom observations by school personnel*. MET project research paper. Harvard University, Cambridge, MA.

Huffcutt, A. (1992). *An empirical investigation of the relationship between multidimensionaldegree of structure and the validity of the employment interview*. Unpublished doctoral dissertation. Texas A&M University, College Station.

Huffcutt, A., & Arthur, W. (1994). Hunter and Hunter (1984) revisited: Interview validity for entry level jobs. *Journal of Applied Psychology*, 79, 184-190.

Huffcutt, A.I., Conway, J.M., Roth, P.L., & Stone, N.J. (2001). Identification and meta-analytic assessment of psychological constructs measured in employment interviews. *Journal of Applied Psychology*, 86(5), 897913.

Jacob, B. & Lefgren, L. (2008). Can principals identify effective teachers? Evidence on subjective performance evaluation in education. *Journal of Labor Economics*, 26, 101-136.

Kansas State Department of Education (2012). Adequate Yearly Progress Fact Sheet. Office of Elementary and Secondary Education. Topeka, KS

Kirchner, J. (2008). *Teacher effectiveness: The relationship of the Teacher Work Sample to the Teacher Insight(TM) interview*. Doctoral dissertation. University of Louisville, Louisville, KY.

Klussman, D. (2004). The impact of teachers selected by the Star Teacher interview process on student achievement, Unpublished dissertation, Seton Hall University, South Orange, NJ.

Koedel, Cory and Julian Betts (2009). Value-Added to What? How a Ceiling in the Testing Instrument Influences Value-Added Estimation. NBER Working Papers 14778, National Bureau of Economic Research, Inc.

Lezotte, L. W. (1991). *Correlates of effective schools: The first and second generation*. Okemos, MI: Effective Schools Products, Ltd.

Longnecker, D. (2006). *An investigation of the metric properties of the Interactive Computer Interview System*. Doctoral dissertation. University of Kansas, Lawrence, KS.

Maurer, S. (2002). A practitioner-based analysis of interview job expertise and scale format as contextual factors in situation interviews. *Personnel Psychology*, 55, 307-327.

Mayfield, E.C. (1964). The selection interview: Re-evaluation of published research. *Personnel Psychology*, 17. 239-260.

McCaffrey, D. F., Lockwood, J. R., Koretz, D. M., & Hamilton, L. S. (2003). Evaluating value-added models for teacher accountability, (MG-158-EDU). Santa Monica, CA: RAND.

McDaniel, M., Whetzel, D., Schmidt F., & Maurer, S. (1994). The validity of employment interviews: A comprehensive review and meta-analysis*. Journal of Applied Psychology, 75*, 322-327.

Medley, Donald M., and Coker, H. (1987). The accuracy of principals' judgments of teacher performance. *Journal of Educational Research 80*, no. 4:242-47.

Metzger, S.A. & Wu, M.J. (2003). Commercial teacher interviews and their problematic role as a teacher qualification. Paper presented at the American Educational Research Association Annual Conference. Chicago, IL: April 2125, 2003.

Metzger, S. A., & Wu, M. J. (2008) Commercial teacher selection instruments: The validity of selecting teachers through beliefs, attitudes, and values. *Review of Education Research, 78*, 4, 921-940.

Miller, J.D., Clements, R.O., & Gardner, C.H. (1977). A preliminary investigation of the teacher perceiver instrument for teacher selection. Austin: University of Texas, Research and Development Center for Teacher Education.

Murnane, Richard (1975). *The impact of school resources on the learning of inner-city children*. Cambridge, MA: Ballinger.

Murray, J. (1990). Review of research on the Myers-Briggs Type Indicator. *Perceptual and Motor Skills*, 70, 1187-1202.

Novotny, M. (2009). *The relationshiop between TeacherInsight scores and professional development and appraisal system domain scores*. Doctoral dissertation. University of North Texas, Denton, TX.

Peterson, Kenneth D. (1987). Teacher evaluation with multiple and variable lines of evidence. *American Educational Research Journal* 24, no. 2:311-17.

Peterson, Kenneth D. (2000). *Teacher evaluation: A comprehensive guide to new directions and practices*. 2nd ed. Thousand Oaks, CA: Corwin.

Pillsbury, P. (2005). Only the best: Hiring outstanding teachers. Leadership, 35(2), 36-38.

Regan, N. (2011). *Accuracy of the TeacherInsight Online Perceiver Tool in Determining the Effectiveness of High Rated and Low Rated Math and Science New Hire Teachers Following One Year and Three Years of Single School District Employment.* Doctoral dissertation. University of Omaha, Omaha, NE.

Reik, M. (2007). *An examination of concurrent validity of the Interactive Computer Interview System (ICIS) using communication arts and math achievement on the Missouri Assessment Program as the Outcome Measure*, Doctoral dissertation, University of Kansas, Lawrence, KS.

Robinson, G. E. (1985). Effective schools research: A guide to school improvement. *Educational research Service, Washington, D.C.*

Rockoff, Johan E. (2004). The impact of individual teachers on student achievement: Evidence from panel data. *American Economic Review* 94, no. 2:247-52.

Sanders, W.L. and Rivers, J.C., (1996). Cumulative and residual effects of teachers on future student academic achievement. Research report, University of Tennessee Value-Added Research and Assessment Center, Knoxville.

Smith, V. (2006). *The development and field testing of a school psychologist employment interview instrument*. Doctoral dissertation. University of Kansas, Lawrence, KS.

Stevenson, G. (2007). A study comparing scores of experienced and inexperienced teachers using the American Association of School Personnel Administrator's Interactive Computer Interview System, Unpublished dissertation, University of Kansas, Lawrence, KS.

Strauss, R. P., & Sawyer, E. A. (1986). Some new evidence on teacher and student competencies. *Economics of Education Review, 5,* 41-48.

Taylor, P., & Small, B. (2002). Asking applicants what they would do versus what they did do: A meta-analytic comparison of situational and past behavior employment interview questions. *Journal of Occupational and Organizational Psychology, 75*, 277-294.

Thayer, C.W. (1978). Project empathy: An alternative way to hire teachers. *North Central Association Quarterly*, 52, 438-442.

The Haberman Foundation. (2009) The Haberman Educational Foundation, Inc. Retrieved April 25, 2009, http://www.altcert.org/Documents/TheHabermanFoundation Brochure.pdf

The New Teacher Project. (2009) Our services. Retrieved April 25, 2009 from http://www.tntp.org/services/teacher_recruitment.html#Selection

U.S. Department of Education (2002). No child left behind: A desktop reference. Office of Elementary and Secondary Education. Washington, D.C.

Wagner, R. (1949). The employment interview: A critical review. *Personnel Psychology,2* 17-46.

Wallwey, D. (2002). *TeacherInsight Interview: Interview development study technical report*. Princeton, NJ: Gallup Organization.

Weishaar, M. (2007). A study of the predictive validity of the Interactive Computer Interview System (ICIS), Unpublished Dissertation, University of Kansas, Lawrence, KS.

Whetzel, D., Baranowski, L., Petro, J., Curtin, P., & Fisher, J. (2003). A written structured interview by any other name is still a selection instrument. *Applied H.R.M. Research 8*(1), 1-16.

Wiesner, W., & Cronshaw, S. (1998). A meta-analytic investigation of the impact of interview format and degree of structure on the validity of the employment interview. *Journal of Occupational Psychology*, 61, 275-290.

Wright, P., Lichtenfels, P., & Pursell, E. (1989). The structured interview: Additional studies and a meta-analysis. *Journal of Occupational Psychology*, 62, 191-199.

Wright, S. P., Horn, S. P., & Sanders, W.L. (1997). Teacher and classroom context effects on student achievement: Implications for teacher evaluation. *Journal of Personnel Evaluation in Education, (11),* 57–67. Retrieved October 15, 2012. http://www.sas.com/govedu/edu/teacher_evalue addedl.pdf.

Appendix

Table A1: Knowledge and Skills identified in Teacher of the Future

Critical Knowledge Needed by Teachers
1. Know the subject(s) they teach and how they are related to other subjects.
2. Know how to teach the subject(s) to students.
3. Know how to teach the subject(s) to students.
4. Know how to plan lessons in a logical sequence.
5. Know how to reflect on their teaching and devise ways of improving it on an ongoing basis.
6. Know how to collaborate with other educators to create the most complete educational environment possible for students.
7. Know how to use technology available to us today, at an intermediate level minimally.
8. Know and appreciate various cultures, and the larger global society and how to establish rapport with a diverse population of students and parents.
9. Know how and where to get needed information and how to educate students to seek and evaluate information.

Critical Skills Needed by Teachers
1. Ability to recognize and respond to individual differences in students.
2. Ability to implement a variety of teaching methods that result in high student achievement.
3. Ability to work cooperatively with parents, colleagues, support staff and supervisors.
4. Ability to display genuine love of teaching students (enthusiasm).
5. Ability to implement full inclusion techniques for special education students.
6. Ability to differentiate instruction for variety of developmental stages and ability levels.
7. Ability to write, speak and present well.
8. Ability to develop critical thinking skills with students.
9. Ability and willingness to relate parents and other community members, individual and corporate, in a positive and helpful fashion.
10. Ability to know and utilize technology in the teaching and learning process.
11. Ability to implement conflict resolution strategies for both adults and students.

Table A2: Assessment Criteria and Domains of Praxis III Educational Testing Service

Domain A: Organizing Content Knowledge for Student Learning
1. Becoming familiar with relevant aspects of students' background knowledge and experiences.
2. Articulating clear learning goals for the lessons that are appropriate to the students.
3. Demonstrating an understanding of the connection between the content that was learned previously, the current content, and the content that remains to be learned in the future.
4. Creating or selecting teaching methods, learning activities, and instructional materials or other resources that are appropriate to the students and that are aligned with the goals of the lesson.
5. Creating or selecting evaluation strategies that are appropriate with the goals of the lesson.

Domain B: Creating an Environment for Student Learning
1. Creating a climate that promotes fairness.
2. Establishing and maintaining rapport with students.
3. Communicating challenging learning.
4. Establishing and maintaining consistent standards of classroom behavior.
5. Making the physical environment as safe and conducive to learning as possible.

Domain C: Teaching for Student Learning
1. Making learning goals and instructional procedures clear to students.
2. Making content comprehensible to students.
3. Encouraging students to extend their thinking.
4. Encouraging students' understanding of content through a variety of means, providing feedback to students to assist learning, and adjusting learning activities as situation demands.
5. Using instructional time effectively.

Domain D: Teacher Professionalism
1. Reflecting on the extent to which the learning goals were met.
2. Demonstrating a sense of efficacy.
3. Building professional relationships with colleagues to share teaching insights and to coordinate learning activities for students.
4. Communicating with parents or guardians about student learning.

Table A3: Framework for Levels of Development within Four Clusters

Cluster	Level of Development
Working with Others	**Level 1:** Egocentric orientation—concerned more about self than others. Others are valued for what they provide. Does not believe in the "social capital" principle where the construction of an interactive web of relationships is important. **Level 2:** Focuses on own classroom but sees the importance of school coordination and interactions with others for the "good of the school". **Level 3:** Altruistic motivation is the driving force for these teachers. Concerned with the larger good. Great respect for "social capital" idea.
Knowledge of Content	**Level 1:** Minimal knowledge. Lacking basic college coursework in much of the field. **Level 2:** Adequate knowledge base typical of a 36 hour college major in the subject field. **Level 3:** Expert knowledge typical of a major with more than 50 hours in the field.
Knowledge of Instruction (Delivery)	**Level 1:** Coherence of an instructional delivery plan is typically lacking. **Level 2:** Teachers consider multipart segments of the lesson, how they fit together, why the sequence is important. **Level 3:** Teacher considers multipart behaviors that are selected and executed based on ongoing analysis of classroom events.
Knowledge of Instruction (Planning)	**Level 1:** Teacher planning behaviors focus on his or her role in the class. The primary focus is on what the teacher intends to do within the class setting relatively independent of considerations about the students' learning goals. **Level 2:** The teacher begins to think about what he or she wants the students to be able to do at defined points in the lesson. **Level 3:** Teacher incorporates branching designs into the planning of their lessons such that they could easily vary the content and method based on classroom feedback.
Knowledge of Instruction (Climate)	**Level 1:** General lack of student attentiveness to academic tasks. **Level 2:** Students are attentive to teacher directed instruction but not necessarily when working in unsupervised groups. **Level 3:** Students self-regulate behavior commensurate with the learning goals.

Knowledge of Instruction (Assessment)	**Level 1:** Assessments viewed as a means of student control. Assessment are infrequent, rarely measure important concepts, poorly designed, with results rarely affecting instruction. **Level 2:** Assessments are viewed as a means of grading and to some extend providing feedback to the teacher about instruction. **Level 3:** Assessments are viewed as a means of diagnosing individual student process and product understanding.
Knowledge of Teaching (Interactions)	**Level 1:** Information and interactions are often confusing for students. They have difficult knowing what is expected. **Level 2:** Teachers engage in typical interchanges with students. Some questions are answered correctly and some incorrectly which usually prompts the teacher to rephrase or return with another question. **Level 3:** Teacher presents information in a way that increases the chances students will comprehend. Teacher thematically connects statements and links student responses to prior material.
Knowledge of Students	**Level 1:** Minimal teacher knowledge of educational psychology. Teacher lacking basic exposure to students of this age or background. **Level 2:** Academic knowledge, student teaching experience, and non-school related teaching exposure to students such as summer camps or church school. **Level 3:** Academic knowledge, teaching experiences in the same context, community experience, and out-of-class contact in students' environment.

Table A4: Demographic Description of Sample Subjects

Assigned Number	Gender	Grade Taught	% Free/Reduced 2010	% Free/Reduced 2011	% Free/Reduced 2012	% Free/Reduced AVERAGE
1	M	3	10.53%	13.04%	22.22%	15.26%
2	F	3	0.00%	4.17%	4.76%	2.98%
3	F	5	63.16%	45.83%	64.71%	57.90%
4	F	5	29.41%	30.00%	18.18%	25.86%
5	M	5	88.24%	72.73%	88.00%	82.99%
6	F	5	9.09%	37.50%	26.92%	24.50%
7	F	5	4.17%	4.00%	3.57%	3.91%
8	F	4	18.18%	10.53%	5.26%	11.32%
9	F	3	0.00%	4.17%	0.00%	1.39%
10	F	3	0.00%	4.17%	4.76%	2.98%
11	F	3	16.67%	14.29%	16.00%	15.65%
12	F	5	47.37%	59.09%	64.71%	57.06%
13	F	3	5.56%	13.64%	23.08%	14.09%
14	F	3	5.88%	0.00%	15.38%	7.09%
15	F	4	78.57%	84.21%	82.35%	81.71%
16	F	4	11.11%	4.76%	10.53%	8.80%
17	F	3	9.09%	8.70%	16.67%	11.48%
18	F	4	55.00%	57.14%	69.57%	60.57%
19	F	5	30.43%	22.22%	27.27%	26.64%
20	F	3	0.00%	4.55%	9.09%	4.55%
21	F	4	9.09%	24.00%	40.00%	24.36%
22	F	4	5.26%	10.71%	12.50%	9.49%
23	F	4	16.67%	9.09%	17.86%	14.54%
24	F	3	18.18%	27.78%	15.79%	20.58%
25	F	3	9.52%	12.00%	0.00%	7.17%
26	F	3	82.35%	75.00%	81.82%	79.72%
27	F	3	16.00%	28.57%	12.00%	18.86%
28	F	4	0.00%	6.90%	0.00%	2.30%
29	F	5	4.55%	4.00%	0.00%	2.85%
30	F	3	4.35%	0.00%	13.04%	5.80%
31	F	4	4.17%	7.41%	0.00%	3.86%
32	F	5	11.11%	18.52%	21.05%	16.89%
33	F	4	25.00%	11.11%	12.50%	16.20%
34	F	3	85.71%	52.94%	72.22%	70.29%
35	F	3	59.09%	47.06%	52.63%	52.93%
36	F	3	5.56%	26.67%	12.50%	14.91%
37	F	5	47.37%	23.81%	18.18%	29.79%
38	F	3	18.18%	5.88%	15.79%	13.28%
39	F	4	4.76%	4.00%	4.17%	4.31%
40	F	5	29.41%	25.00%	20.83%	25.08%

Table A5: Teacher ICIS Interview Scores with Residual Reading and Residual Math Adjustments

Assigned Number	Theme 1	Theme 2	Theme 3	Theme 4	Total	Residual Reading	Residual Math
1	2.00	2.00	2.83	2.40	2.37	0.90	0.59
2	2.40	2.75	2.50	2.60	2.54	-0.46	0.05
3	3.00	2.75	2.83	2.75	2.82	-0.50	-0.58
4	3.00	3.00	2.83	3.00	2.93	0.10	0.90
5	1.75	1.60	2.17	2.75	2.05	0.73	0.46
6	3.00	2.60	2.83	2.75	2.78	-0.61	-0.54
7	3.00	2.60	2.83	2.60	2.74	-0.03	-0.47
8	3.00	3.00	3.00	3.00	3.00	0.09	-0.56
9	2.20	1.80	2.40	2.25	2.21	-0.53	-0.67
10	3.00	3.00	2.67	2.75	2.81	-0.50	-0.07
11	2.25	2.60	2.50	2.25	2.42	0.74	0.65
12	2.25	2.00	2.83	2.00	2.33	-0.50	-0.72
13	2.00	1.40	2.33	2.00	1.94	0.43	0.80
14	2.25	1.80	2.67	2.20	2.25	-0.20	-0.40
15	2.75	2.75	2.83	3.00	2.82	-0.19	0.63
16	2.25	2.60	2.83	2.75	2.63	-0.21	-0.20
17	2.60	2.40	2.83	2.40	2.57	-1.46	-1.55
18	3.00	2.00	2.83	2.60	2.65	0.67	-0.01
19	2.20	2.40	2.83	2.40	2.48	1.72	1.44
20	2.00	2.75	2.67	3.00	2.63	0.25	0.17
21	2.60	3.00	2.50	2.40	2.58	-0.03	0.58
22	3.00	3.00	2.67	3.00	2.87	-0.87	0.06
23	2.40	2.75	3.00	3.00	2.78	-0.25	-1.71
24	2.00	2.40	2.33	2.25	2.25	0.55	0.87
25	3.00	3.00	3.00	3.00	3.00	0.05	0.00
26	3.00	3.00	2.83	2.75	2.88	-0.50	-0.92
27	2.60	2.40	3.00	2.75	2.70	0.81	0.83
28	2.75	2.75	3.00	3.00	2.88	-1.19	-1.48
29	2.60	3.00	3.00	3.00	2.88	-0.38	-0.88
30	2.40	2.40	3.00	3.00	2.68	-0.19	-0.66
31	2.75	2.75	3.00	3.00	2.88	-0.32	-0.29
32	2.75	2.25	2.83	2.40	2.58	0.87	0.37
33	2.75	3.00	2.67	2.75	2.76	-0.20	0.96
34	2.20	2.00	3.00	3.00	2.59	-0.78	-0.94
35	1.75	2.60	2.10	2.75	2.26	-0.58	-0.42
36	3.00	2.60	3.00	2.75	2.83	-0.01	-0.28
37	3.00	2.40	2.67	2.60	2.63	0.56	1.19
38	2.60	2.60	2.67	3.00	2.68	-0.46	0.40
39	3.00	3.00	2.83	3.00	2.93	1.85	1.50
40	2.25	2.75	2.80	3.00	2.71	0.60	0.91

Table A6: ICIS Total Weighted Average and Administrator Rating

Assigned Number	ICIS Total	Administrator Rating
1	2.37	8
2	2.54	8
3	2.82	7
4	2.93	10
5	2.05	9
6	2.78	9
7	2.74	6
8	3.00	9
9	2.21	9
10	2.81	8
11	2.42	9
12	2.33	6
13	1.94	10
14	2.25	6
15	2.82	10
16	2.63	7
17	2.57	3
18	2.65	7
19	2.48	8
20	2.63	7
21	2.58	8
22	2.87	6
23	2.78	9
24	2.25	6
25	3.00	9
26	2.88	8
27	2.70	8
28	2.88	9
29	2.88	9
30	2.68	6
31	2.88	9
32	2.58	8
33	2.76	10
34	2.59	6
35	2.26	5
36	2.83	10
37	2.63	10
38	2.68	9
39	2.93	10
40	2.71	10

Table A7: ICIS Total Weighted Average Scores with Administrator Rating, Residual Reading, and Residual Math

Assigned Number	ICIS Total	Administrator Rating	Residual Reading	Residual Math
1	2.37	8	0.90	0.59
2	2.54	8	-0.46	0.05
3	2.82	7	-0.50	-0.58
4	2.93	10	0.10	0.90
5	2.05	9	0.73	0.46
6	2.78	9	-0.61	-0.54
7	2.74	6	-0.03	-0.47
8	3.00	9	0.09	-0.56
9	2.21	9	-0.53	-0.67
10	2.81	8	-0.50	-0.07
11	2.42	9	0.74	0.65
12	2.33	6	-0.50	-0.72
13	1.94	10	0.43	0.80
14	2.25	6	-0.20	-0.40
15	2.82	10	-0.19	0.63
16	2.63	7	-0.21	-0.20
17	2.57	3	-1.46	-1.55
18	2.65	7	0.67	-0.01
19	2.48	8	1.72	1.44
20	2.63	7	0.25	0.17
21	2.58	8	-0.03	0.58
22	2.87	6	-0.87	0.06
23	2.78	9	-0.25	-1.71
24	2.25	6	0.55	0.87
25	3.00	9	0.05	0.00
26	2.88	8	-0.50	-0.92
27	2.70	8	0.81	0.83
28	2.88	9	-1.19	-1.48
29	2.88	9	-0.38	-0.88
30	2.68	6	-0.19	-0.66
31	2.88	9	-0.32	-0.29
32	2.58	8	0.87	0.37
33	2.76	10	-0.20	0.96
34	2.59	6	-0.78	-0.94
35	2.26	5	-0.58	-0.42
36	2.83	10	-0.01	-0.28
37	2.63	10	0.56	1.19
38	2.68	9	-0.46	0.40
39	2.93	10	1.85	1.50
40	2.71	10	0.60	0.91

Table A8: Kansas Assessment Mean Scores, Z Scores and Residual Scores for Reading and Math

Assigned Number	Grade	Class Avg. Reading	Reading Z Score	Residual Reading Score	Class Avg. Math	Math Z Score	Residual Math Score
1	3	86.70	1.14	0.90	91.03	1.27	0.59
2	3	83.20	-0.11	-0.46	90.23	1.08	0.05
3	5	80.43	-1.09	-0.50	78.67	-1.67	-0.58
4	5	84.47	0.34	0.10	87.90	0.53	0.90
5	5	81.87	-0.58	0.73	80.67	-1.19	0.46
6	5	82.67	-0.30	-0.61	81.97	-0.88	-0.54
7	5	85.80	0.82	-0.03	84.23	-0.34	-0.47
8	4	84.37	0.31	0.09	81.60	-0.97	-0.56
9	3	83.10	-0.15	-0.53	88.17	0.59	-0.67
10	3	83.07	-0.16	-0.50	89.87	0.99	-0.07
11	3	86.10	0.92	0.74	91.20	1.31	0.65
12	5	80.50	-1.07	-0.50	78.17	-1.79	-0.72
13	3	85.20	0.60	0.43	91.77	1.45	0.80
14	3	83.70	0.07	-0.20	88.60	0.69	-0.40
15	4	81.73	-0.63	-0.19	80.80	-1.16	0.63
16	4	84.10	0.21	-0.21	82.43	-0.77	-0.20
17	3	78.93	-1.63	-1.46	84.80	-0.21	-1.55
18	4	83.43	-0.03	0.67	80.47	-1.24	-0.01
19	5	88.80	1.88	1.72	90.07	1.04	1.44
20	3	85.50	0.71	0.25	90.50	1.14	0.17
21	4	83.80	0.10	-0.03	83.30	-0.57	0.58
22	4	83.30	-0.07	-0.87	82.93	-0.65	0.06
23	4	83.87	0.13	-0.25	79.13	-1.56	-1.71
24	3	84.97	0.52	0.55	91.53	1.39	0.87
25	3	84.53	0.36	0.05	89.80	0.98	0.00
26	3	75.63	-2.80	-0.50	82.07	-0.86	-0.92
27	3	86.03	0.90	0.81	91.53	1.39	0.83
28	4	83.17	-0.12	-1.19	80.13	-1.32	-1.48
29	5	84.93	0.51	-0.38	82.60	-0.73	-0.88
30	3	83.83	0.12	-0.19	87.90	0.53	-0.66
31	4	84.13	0.22	-0.32	82.47	-0.76	-0.29
32	5	87.23	1.32	0.87	86.50	0.19	0.37
33	4	83.87	0.13	-0.20	84.43	-0.30	0.96
34	3	75.60	-2.81	-0.78	82.67	-0.72	-0.94
35	3	77.97	-1.97	-0.58	85.40	-0.07	-0.42
36	3	83.60	0.03	-0.01	88.43	0.65	-0.28
37	5	85.43	0.68	0.56	88.73	0.72	1.19
38	3	82.20	-0.47	-0.46	90.60	1.17	0.40
39	4	86.67	1.12	1.85	86.07	0.09	1.50
40	5	85.90	0.85	0.60	88.00	0.55	0.91

Table A9: Correlations of ICIS Interview Sub-Scores, Residual Reading and Residual Math

		Working with Others	Knowledge of Content	Knowledge of Teaching	Knowledge of Students	ICIS Total Weighted Average Score	Residual Reading	Residual Math
Working with Others	Pearson Correlation							
	Sig. (2-Tailed)							
Knowledge of Content	Pearson Correlation	.577**						
	Sig. (2-Tailed)	.000						
Knowledge of Teaching	Pearson Correlation	.560**	.346*					
	Sig. (2-Tailed)	.000	.029					
Knowledge of Students	Pearson Correlation	.378*	.624**	.464**				
	Sig. (2-Tailed)	0.16	.000	.003				
ICIS Total Weighted Average Score	Pearson Correlation	0.813**	.828**	.741**	.753**			
	Sig. (2-Tailed)	.000	.000	.000	.000			
Residual Reading	Pearson Correlation	-.144	-.171	-.087	-.172	-.177		
	Sig. (2-Tailed)	.375	.291	.592	.288	.276		
Residual Math	Pearson Correlation	-.074	-.014	-.318*	-.140	-.163	.770**	
	Sig. (2-Tailed)	.652	.933	.045	.387	.315	.000	

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

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

Table A10: Quartile Values - Reading

Grade	Year	Minimum	Maximum	First Quartile	Second Quartile	Third Quartile
3	2010	0.86	-1.82	-0.84	-0.18	0.28
3	2011	1.28	-1.02	-0.44	0.06	0.71
3	2012	0.86	-1.54	-0.59	0.19	0.58
4	2010	2.00	-1.26	-0.71	-0.64	0.09
4	2011	1.26	-0.90	-0.60	0.47	1.01
4	2012	1.43	-1.83	-0.81	-0.28	0.73
5	2010	1.76	-0.97	-0.69	-0.18	1.15
5	2011	1.55	-0.66	-0.08	0.39	0.77
5	2012	1.99	-1.13	-0.72	-0.08	0.72

Table A11: Quartile Values - Math

Grade	Year	Minimum	Maximum	First Quartile	Second Quartile	Third Quartile
3	2010	0.71	-1.97	-0.99	-0.11	0.32
3	2011	1.47	-1.36	-0.19	0.07	0.63
3	2012	1.25	-1.68	-0.76	0.21	0.48
4	2010	1.77	-1.20	-1.03	-0.25	0.39
4	2011	1.22	-1.70	-0.51	-0.22	0.57
4	2012	2.39	-1.95	-1.13	0.05	1.06
5	2010	1.20	-1.59	-1.09	-0.32	0.87
5	2011	1.36	-1.45	0.00	0.39	1.03
5	2012	1.65	-1.00	-0.75	-0.05	1.01

Table A12: Residual Scores by Grade and Year

Assigned Number	Grade	Reading 2010	Reading 2011	Reading 2012	Math 2010	Math 2011	Math 2012
1	3	0.26	1.28	0.86	0.46	0.70	0.34
2	3	-0.20	-0.44	-0.45	-0.09	-0.07	0.33
9	3	-0.33	-0.44	-0.51	-0.96	-0.63	-0.07
10	3	-0.20	-0.41	-0.58	-0.21	-0.07	0.13
11	3	0.86	0.78	0.34	0.71	0.53	0.42
13	3	0.58	0.69	-0.08	0.48	0.60	0.97
14	3	0.31	0.38	-1.12	-0.26	0.49	-1.20
17	3	-1.39	-1.02	-1.34	-1.85	-0.38	-1.68
20	3	-0.36	0.33	0.79	-0.12	0.29	0.29
24	3	0.11	0.78	0.56	0.27	1.47	0.47
25	3	0.09	-0.23	0.35	0.00	0.10	-0.07
26	3	-0.81	-0.93	0.14	-0.03	-1.36	-1.04
27	3	0.54	0.98	0.63	0.23	1.07	0.82
30	3	-0.16	-0.51	0.24	-0.85	-0.59	-0.19
34	3	-1.49	-0.14	-0.64	-1.11	-0.12	-1.21
35	3	-1.82	-0.27	0.44	-1.54	-0.09	0.51
36	3	-0.91	0.25	0.69	-1.97	0.03	1.25
38	3	0.13	0.26	-1.54	0.63	1.07	-0.67
8	4	-0.16	0.47	-0.28	-1.03	0.36	-0.68
15	4	0.24	1.06	-1.00	1.77	0.15	-0.32
16	4	-0.03	1.26	-1.83	-0.14	0.76	-1.13
18	4	-0.64	1.01	1.41	-0.19	-0.51	0.74
21	4	0.09	0.55	-0.70	-0.25	0.57	1.06
22	4	-0.68	-0.44	-0.81	-0.36	-0.22	0.71
23	4	-0.71	-0.60	0.69	-1.13	-1.70	-1.23
28	4	-1.26	-0.65	-0.75	-1.20	-0.41	-1.95
31	4	-0.71	-0.41	0.23	-0.46	-0.32	0.05
33	4	-1.04	-0.90	1.43	0.39	-0.53	2.39
39	4	2.00	0.78	0.73	0.92	1.22	1.38
3	5	-0.58	-0.66	-0.08	-0.36	-1.45	-0.05
4	5	0.11	0.77	-0.64	0.87	1.08	0.53
5	5	1.76	0.72	-0.75	0.87	0.74	-0.77
6	5	-0.69	-0.51	-0.32	-1.09	-0.15	-0.28
7	5	-0.78	0.33	0.42	-0.67	0.14	-0.62
12	5	-0.97	0.78	-1.13	-1.25	0.00	-1.00
19	5	1.15	1.55	1.61	0.96	1.36	1.65
29	5	-0.52	0.33	-0.72	-1.59	0.05	-0.75
32	5	-0.18	0.39	1.99	-0.32	0.39	1.01
37	5	1.26	-0.08	0.22	1.20	1.01	1.04
40	5	0.11	0.69	0.72	0.87	1.03	0.60

Table A13: Quartile Ranking by Grade and Year

Assigned Number	Grade	Reading 2010	Reading 2011	Reading 2012	Math 2010	Math 2011	Math 2012
1	3	3	4	4	4	4	3
2	3	2	1	2	3	2	3
9	3	2	1	2	2	1	2
10	3	2	2	2	2	2	2
11	3	4	4	3	4	3	3
13	3	4	3	2	4	3	4
14	3	4	3	1	2	3	1
17	3	1	1	1	1	1	1
20	3	2	3	4	2	3	3
24	3	3	4	3	3	4	3
25	3	3	2	3	3	3	2
26	3	2	1	2	3	1	1
27	3	4	4	4	3	4	4
30	3	3	1	3	2	1	2
34	3	1	2	1	1	2	1
35	3	1	2	3	1	2	4
36	3	1	3	4	1	2	4
38	3	3	3	1	4	4	2
8	4	3	2	2	1	3	2
15	4	4	4	1	4	3	2
16	4	3	4	1	3	4	1
18	4	2	3	4	3	1	3
21	4	3	3	2	2	3	3
22	4	2	2	1	2	2	3
23	4	1	1	3	1	1	1
28	4	1	1	2	1	2	1
31	4	2	2	3	2	2	2
33	4	1	1	4	3	1	4
39	4	4	3	3	4	4	4
3	5	2	1	2	2	1	2
4	5	3	3	2	3	4	3
5	5	4	3	1	3	3	1
6	5	1	1	2	1	1	2
7	5	1	2	3	2	2	2
12	5	1	4	1	1	1	1
19	5	3	4	4	4	4	4
29	5	2	2	1	1	2	1
32	5	2	2	4	2	2	3
37	5	4	1	3	4	3	4
40	5	3	3	3	3	3	3

Table A14: Quartile Variance - Reading

Assigned Number	Grade	Reading 2010	Reading 2011	Reading 2012	Variance
10	3	2	2	2	0.00
17	3	1	1	1	0.00
40	5	3	3	3	0.00
27	3	4	4	4	0.00
19	5	3	4	4	0.33
1	3	3	4	4	0.33
24	3	3	4	3	0.33
39	4	4	3	3	0.33
11	3	4	4	3	0.33
3	5	2	1	2	0.33
9	3	2	1	2	0.33
29	5	2	2	1	0.33
22	4	2	2	1	0.33
2	3	2	1	2	0.33
26	3	2	1	2	0.33
6	5	1	1	2	0.33
28	4	1	1	2	0.33
34	3	1	2	1	0.33
31	4	2	2	3	0.33
21	4	3	3	2	0.33
25	3	3	2	3	0.33
4	5	3	3	2	0.33
8	4	3	2	2	0.33
7	5	1	2	3	1.00
20	3	2	3	4	1.00
13	3	4	3	2	1.00
18	4	2	3	4	1.00
35	3	1	2	3	1.00
23	4	1	1	3	1.33
32	5	2	2	4	1.33
30	3	3	1	3	1.33
38	3	3	3	1	1.33
37	5	4	1	3	2.33
14	3	4	3	1	2.33
5	5	4	3	1	2.33
16	4	3	4	1	2.33
36	3	1	3	4	2.33
12	5	1	4	1	3.00
15	4	4	4	1	3.00
33	4	1	1	4	3.00

Table A15: Quartile Variance - Math

Assigned Number	Grade	Math 2010	Math 2011	Math 2012	Variance
10	3	2	2	2	0.00
17	3	1	1	1	0.00
40	5	3	3	3	0.00
19	5	4	4	4	0.00
39	4	4	4	4	0.00
31	4	2	2	2	0.00
7	5	2	2	2	0.00
23	4	1	1	1	0.00
12	5	1	1	1	0.00
27	3	3	4	4	0.33
1	3	4	4	3	0.33
24	3	3	4	3	0.33
11	3	4	3	3	0.33
4	5	3	4	3	0.33
13	3	4	3	4	0.33
37	5	4	3	4	0.33
3	5	2	1	2	0.33
9	3	2	1	2	0.33
30	3	2	1	2	0.33
29	5	1	2	1	0.33
6	5	1	1	2	0.33
28	4	1	2	1	0.33
34	3	1	2	1	0.33
22	4	2	2	3	0.33
2	3	3	2	3	0.33
21	4	2	3	3	0.33
25	3	3	3	2	0.33
20	3	2	3	3	0.33
32	5	2	2	3	0.33
8	4	1	3	2	1.00
14	3	2	3	1	1.00
15	4	4	3	2	1.00
38	3	4	4	2	1.33
26	3	3	1	1	1.33
18	4	3	1	3	1.33
5	5	3	3	1	1.33
35	3	1	2	4	2.33
16	4	3	4	1	2.33
36	3	1	2	4	2.33
33	4	3	1	4	2.33

Detailed Description of Interactive Computer Interview System

Delivery of Instruction

According to the scoring rubric of the ICIS, a candidate with the highest level of competency in the area of “Delivery of Instruction” would have the following characteristics:

- Multi-part teacher behaviors that are selected and executed based upon ongoing analysis of classroom events.
- The teacher generally makes future directional decisions based on current classroom events.
- New instructional designs might be devised on the spot.

Additionally, the interviewer should hear the teacher candidate make an intentional shift from a direct presentation to a review when it becomes apparent that certain students do not understand important concepts or processes. Subsequent analysis of the teacher’s part (by additional probes of student understanding) sets a direction for future teaching behaviors. Additionally, the teacher’s behaviors should initiate a class discussion and make decisions about the direction of the lesson from the classroom. The teacher must make active decisions concerning what important concepts to summarize the discussion and where to direct the class next.

Planning

According to the scoring rubric of the ICIS, a candidate with the highest level of competency in the area of “Planning” would begin to incorporate branching designs into

the planning of their lessons so they could easily vary the content and method based upon classroom feedback. Furthermore, the candidate would have several possibilities from which they could pursue, but would wait to make final decisions until they obtain the additional diagnostic feedback from the class or individual students.

Listed below is an ICIS example of what a prospective candidate may use as his/her story to convey their knowledge of being an effective planner: *A chemistry teacher thinks to him/herself - although it is important for the students to understand how the Periodic Table is organized, it is more important they understand the process scientists use to attempt to make sense of seemingly disparate bits of information. To give the students a sense of this discovery process, I will give each group of students 50 cards containing information about 50 different fictitious atoms. It will be there job to organize the 50 cards into some form of order. There are several possibilities the students could come up with (arranging by size, color, state of matter, melting point, etc. or some combination) so I will have to wait until I see their logic before the next part of the lesson can be designed. For example, if they put the cards in order of the number of electrons in the outer shell, then we can talk about families of elements. I have in mind about 10 different ordering concepts, but will discuss them in order of their discovery by students.*

Interactions with Students

According to the scoring rubric of the ICIS, a candidate with the highest level of competency in the area of “Interactions with Students” would have the following characteristics:

- The teacher presents information in a way that increases the chances students will comprehend.
- He/she thematically connects statements and links student responses to prior material.
- The teacher uses vocabulary familiar to students and rephrases when necessary.
- The teacher poses questions that are understandable to students and rephrased when needed for additional clarity.
- He/she uses discourse marker techniques to indicate what is important in the subject matter including marking expressions, repetition, and numeration of major points.
- The teacher employs non-verbal behavior as a way of signaling students and negative student responses are dignified and redirected ultimately searching for an opportunity.

Some examples, according to the ICIS scoring rubric, that an interviewer would want to hear from a candidate in order to effectively convey their understanding of the role “Interactions with Students” has to do with being an effective educator would include: the teacher asks questions using information familiar to the student’s background. For example, the teacher may ask a student, “What are the colors of Germany’s flag?” When the student responds, “Red, white and blue,” the teacher, instead of admonishing the incorrect response, would respond, “I think you are thinking

of the United States flag. Why don't you take a few moments to look that one up and I'll ask you again in a few minutes."

Assessment

According to the scoring rubric of the ICIS, a candidate with the highest level of competency in the area of "Assessment" would have the following characteristics:

- Views assessments as a means of diagnosing individual student process and product understanding.
- Students are carefully prepared to take the assessments.
- Multiple dimensions of student understanding and performance are measured.
- Assessments are well designed and scored.
- Student feedback is rapid, detailed, and addresses student strengths and weaknesses.

Some examples, according to the ICIS, that an interviewer would want to hear from a candidate in order to effectively convey their knowledge of assessments would include: students practice assessments via simulations; feedback is detailed, prompt, and individualized; corrective teaching follows errors, assessments are multi-dimensional and measure various aspects of the lesson - cognitive achievement, skill development, etc.; assessments measure various achievement levels - analysis, synthesis, knowledge acquisition, etc.

Climate Development

According to the scoring rubric of the ICIS, a candidate with the highest level of competency in the area of “Climate Development” would have the following characteristics:

- Students would self-regulate behavior commensurate with the learning goals.
- Discipline problems are rare.
- Students feel comfortable, cohesive, secure, interested, and value learning.
- Students will share feelings and aspirations with the teacher and class members.

Some examples according to the ICIS of what you would expect to hear an interviewee discuss would include: Rules are jointly developed and enforced by the teacher and students; students feel free to express opinions minority opinions - even very radical ones; teacher knows exactly what is going on everywhere and can anticipate almost all events; pace is lively and directed toward learning objectives; teacher praise is appropriated, directed as needed, and functions to support classroom interactions and learning; teacher can leave the room for brief periods with little degradation in learning; enthusiasm is common among the students - a certain “electricity” can be felt within the classroom.