

**A Study of the Construct Validity of the Interactive Computer Interview System
(ICIS) using Student Evaluations as the Outcome Measure**

By

Chad G. Evans

C2010

B.S. Hastings College, 1997

M.S. Northwest Missouri State University, 1999

Submitted to the Department of Educational Leadership and Policy Studies and
the Faculty of the Graduate School of the University of Kansas in partial
fulfillment of the requirements for the degree of Doctor of Education.

Dissertation Committee:

Chairperson

Dissertation defended: January 14, 2010

The Dissertation Committee for Chad Evans certifies
that this is the approved version of the following dissertation:

Chad Gerhold Evans

Committee:

_____ Chairperson*

Date approved: _____

Table of Contents

Abstract.....3

I. Introduction4

II. Review of Literature.....9

 A Teacher’s Environment: Where Does He/She Belong.....14

 Common Problems: What Do Weak Interviews Supply15

 More Subtle Problems: What Throws an Interview Off Track?.....16

 Lack of Structure: What Derives from a Poor Interviews20

 The Idealistic Approach: What if the Situation were Perfect?21

 Conclusion: Using the Right Tools for the Job.....22

III. Method24

 The Student Survey Instrument25

 The ICIS Interview Instrument26

 Structure of the Data Matrix27

 Data Collection Methods27

IV. Results.....29

 Descriptive Statistics.....29

 Concurrent Validity29

 Examination of Aspects of Bias.....32

V. Discussion34

 Limitations of the Current Study36

 Suggestions for Further Research and Improvements38

 Conclusion39

VI. References.....41

VII. Appendices46

 Appendix A: Criteria for Data Collection by External Team of Experts47

 Appendix B: Steps to Excellence Questionnaire48

 Appendix C: ICIS Scoring Rubrics.....50

 Appendix D: Teacher Scores on the ICIS.....51

Abstract

The present study sought to investigate the concurrent validity of the ICIS employment interview tool as measured by student satisfaction by examining the correlation between the scores obtained through interviews of forty high school teachers and the affective responses of their students as measured through a Likert survey instrument. The teachers answered programmed questions in four major categories (Knowledge of Teaching, Working with Others, Knowledge of Content, and Knowledge of Students) in interviews conducted within a three-week period at the conclusion of the school year. These randomly chosen teachers of required freshman and sophomore classes represented teaching experience from one to twenty-eight years and ages from twenty-four to fifty-eight years. All students of the selected classes completed a thirty-question "Steps to Excellence Student Questionnaire" that measured student satisfaction levels. The 1039 student surveys were completed in April, giving the students as much time as possible with the teachers they rated. Pearson calculations showed a statistically significant correlation between the student survey average scores and all four sub-scales measured by the ICIS instrument and the ICIS total rating.

I. Introduction

“Crisis” has become the national watchword for focus and action over the past twenty-five years, ever since the 1983 publication of *A Nation at Risk*. Americans face crises in nearly every major area of their citizenship, and crises in the schools affect all citizens, because everyone has attended school, sent their children to school, or interacted with former students of the U.S. public school system. In response to the crisis in the classroom, officials from a variety of leadership positions (Congress, state legislatures, educational organizations, to name a few) have asserted their assumed authority over the public schools. In the end, however, if the problems in the classroom are to be resolved in any meaningful and lasting way, the emphasis must come in the arena of the crisis: the classroom. As the *No Child Left Behind* informational publication points out, “One of the most important ways to close the achievement gap and provide all children with a great education is to provide them with great teachers. Studies have shown the single greatest school-controlled effect on student achievement is teacher quality” (2004, p. 12). Efforts outside the classroom will have, and have had, little meaningful impact on the quality of teaching and the subsequent improvement of student performance. Every useful examination of student progress, whether conducted by schools of education, congressional committees, or Secretaries of Education, have identified *only one* omnipresent factor in the success or failure of students: the teacher, because “the productive processes of teaching and learning are distinctively dependent on teachers” (Harris, et al., 1989, p. 728). School districts can build new buildings, purchase expensive textbooks and curricula, and provide extensive in-service training, and a host of other additions to their school’s efforts to improve student performance, but nothing can

replace the hiring of good teachers. Even good teachers face challenges that boards of education cannot change: “For teaching to be both good and successful, it must be conjoined with factors well beyond the range of control of the classroom teacher” (Fenstermacher and Richardson, 2005, p. 186). Student preparedness, willingness to learn, socio-economic conditions, and community standards and attitudes enter into the teachers’ decisions and direction every day. Choosing the right teacher to perform under these requirements is a serious task in every school setting.

An administrator also must evaluate candidates with regard to the context of a given school: what selection tool can an administrator effectively apply to compare an adaptive physical education teacher to a teacher of advanced placement physics students? The useful comparability of various teachers, teaching styles, teacher training, and experience becomes even less likely once elements of a pluralistic society enter the hiring equation. Since 2002, the NCLB act has sought to establish *one* standard for *all* schools, regardless of their funding, student population, diversity of the school’s community, or the place of the school or district in the process of long-term planning. McWalters (2004) asks, “Why should an urban school district, with its high percentage of students with limited proficiency in English, its high rate of poverty, and its high dropout rate, get the same amount of funding per pupil as a more affluent suburban district?” (p. 69). He’s clearly pointing out the failure of a one-standard philosophy in terms of funding. The same one-standard philosophy has the same negative effect on hiring in our incredibly diverse population. State and national standards are too broad to accommodate the almost immeasurable diversity of U.S. schools. The reality is that these evaluations every year become snap shots by which schools are evaluated irrespective of the on-going nature of

education; the student's ability at the end of his school career becomes less important than his immediate situation.

Regardless of the number of external criteria imposed upon the schools, the teacher remains the one critical element in satisfying the demands of NCLB or, in fact, any other program of evaluation. The one true failing of all such endeavors to evaluate student success is that measuring the schools means nothing if the examination ends with the overall scores of the school. McWalters (2004) emphasizes that "standards, expectations, testing and accountability do not really matter unless it leads to better teaching, better learning, and improved results" (p. 70). To this end, nearly all states have implemented rigorous measures for students, and by these results school districts receive a "grade card" for the year. So while this or any other device may measure a school's overall performance, it tells nothing about an individual student. Every student and parent can identify his/her favorite teacher or teachers, and while these identifications often include teacher personality, truly memorable teaching involves student achievement; a student feels good about a teacher when he/she feels good about him/herself, and a student feels good about him/herself when he/she feels the positive experience of learning something new or developing a new skill, and researchers have found that even "pupils at the elementary level can perceive a difference between effective and ineffective teachers" (Manning and Payne, 1984, p. 10).

The strength of any school is its teachers, and the strength of an administrator is his/her ability to identify and hire the best available teachers. Typically, this process has included using a review of references and transcripts to identify teachers to interview, with the interview serving as the final element in the hiring decision. Research shows that

while many administrators believe they have developed a special sense for identifying good teachers, interviewer biases (though unintended) and practices (often more comfortable than effective) can generate faulty information and impressions, leading to a bad hire (Harris, 1989, p. 696). Attempts have been made to rectify this situation by using a pre-selected set of questions for all teacher applicants. The attendant challenge of recording and remembering responses was next assisted by developing rubrics to assess the answers. Question sequence structuring (attempting to build on each response with a related question) followed the developing of the rubrics. The final element in effective interviewing came with the “branching questions” (Tengler and Jablin, 1983, p. 247), questions triggered by applicants’ responses. The benefit of all these additions to basic questioning, however, suffered from having too much material to keep track of, often interfering with the flow of the interview.

The use of computer technology, however, has sought to change the nature and effectiveness of the structured interview by greatly reducing the recording and remembering aspects and by automatically directing the interviewer to branching questions. The computer programs can even redirect the branching questions back to the main line of questioning for the interviewer. Several such programs/instruments have been developed.

The present paper sought to conduct a concurrent validity study of the ICIS computer assisted instrument by comparing students’ evaluations (“Steps to Excellence” Student Questionnaire [see Appendix B]) of their teachers to the scores derived from the ICIS employment interview system (ICIS [see Appendix C]). The study sought to build on the research of other studies that have evaluated the effectiveness of the ICIS interview

system *vis a vis* standardized testing (Reik, 2007), and principal subjective evaluations (Ebmeier and Ng, 2004).

II. Review of Literature

The foundation of the current study comes from an examination of published literature in the field of personnel matters in general and teacher hiring practices in particular. Professional journals and university studies over the past twenty years provided the source material for the study. These documented research efforts evaluated the results of structured and unstructured interviews to determine the relative effectiveness of each approach in order to provide recommendations for the improvement of hiring experiences. Additional literature describing quality teaching and quality teachers helped to define the goals of the interview instruments examined.

Within this stated context of the current review, one could assert that the best schools integrate the performance of every member of the staff, student body, and community to create the best possible learning experience for the students and, by extension, the brightest future for the community. While students, the primary subject of all school research, interact at some time with virtually all members of the staff, their most consistent, compelling, and long term relationship remains, and will remain, the relationship with their teachers. And because each teacher generally has just one extended opportunity to impact the students' futures, each school must make the most of its opportunities to provide the best possible teacher: "Excellent teachers, those able to contribute for the long haul, must know: the student and his/her individual nuances, the subject matter and how to teach the subject matter, and a variety of instructional strategies" (Ryan and Alcock, 2002, p. 62). Such excellent teachers do not come as a response to the NCLB guidelines, because "to be 'highly qualified,' according to NCLB, a teacher must only meet three criteria: He or she must (a) have a bachelor's degree; (b)

be fully certified or licensed, including certification obtained through ‘alternative routes’ but excluding licensure that has been ‘waived on an emergency, temporary, or provisional basis’; and (c) have demonstrated content knowledge in the subject her or she teaches” (Smith, Desimone, and Ueno, 2005, p. 88). The “highly qualified” teacher can clearly be identified by an application form and test results, but identifying the “excellent” teacher requires much more than a data sheet and a transcript. And since very few administrators have the luxury or opportunity to observe prospective teachers in the classroom, the administrator’s task is to find a way to identify those “excellent” qualities using an employment interview. He must do this in the face of the most incongruous of public attitudes: “In many parts of the United States, one commonly hears that the brightest college students would be ill advised to waste their capacities as public school teachers. This disparagement has always seemed blasphemous to educators in view of the overwhelming importance of the development of our young people’s capacities and in view of the inherent majesty of a profession dedicated to that development. But pleas for higher status for teachers have fallen on deaf ears” (Renfield, 1969, p. 94).

Modern administrators face increased responsibilities and duties with each successive year’s employment. Given the immense and diverse challenges engendered by local, state, and federal mandates, the job of recruiting and hiring new teachers can take on a mechanical form with much of the hiring determined by quotas, guidelines, and requirements that may have little recognition of the individual needs of any one particular school. Administrators must focus on finding the right teacher for the position and hold the legislated mandates in the background. According to Michael W. Jinks, Superintendent of the Ballard Community School District: “The greatest causes of weak

interviews are related to problems with questioning skills and preparation” (Caldwell, 1993, p. 41). Additionally, “interviewer cognitive complexity may increase the accuracy of ratings, whereas interviewer mood may decrease accuracy” (Posthuma, Morgeson, and Campion, 2002, p. 28). In other words, the quality of the interview may be impacted by factors unintended and unexpected by the interviewer. These challenges of preparation and mood (within the complexity of guiding a modern public school) call for a method of interview that offers consistency in an otherwise inconsistent interviewing environment. As Caldwell (1993) concludes: “A structured interview would be more reliable because it would force the interviewer to be more attentive” (p. 24). Arvey (1995) further explains: “Structure may reduce information processing requirements and overload, thus allowing interviewers to focus on candidate responses.” In the organic, fluid environment of a modern school, any appropriate structure becomes of paramount importance.

As stated in “A Review of Structure in the Selection Interview” (Campion, et al., 1997), “In the 80-year history of published research on employment interviewing (dating back to Scott, 1915), few conclusions have been more widely supported than the idea that structuring the interview enhances reliability and validity” (p. 655). Compare this with concurrent findings that personnel professionals continue to rely more on their “experience” and beliefs over demonstrated facts and statistics.

There was a controversy in the 1950s whether statistical methods of combining data were better than clinical methods where expert judgment was used. Interviewing was embroiled in this controversy. The evidence from many contexts overwhelmingly favors statistical prediction (Campion, et al., 1997, p. 688).

Further, “Meta-analytic reviews of validity studies have also unanimously supported the superiority of structured interviews” (Campion, 1997, et al., p. 656).

However, personnel professionals often favor details not supported by actual research findings. According to Ryan and Tippins (2004), wide gaps between knowledge and practice exist in many organizations (p. 306). For example, Rynes, Colbert, and Brown (2002), reported that “72% of the HR managers they surveyed thought that, on average, conscientiousness is a better predictor of employee performance than intelligence, whereas the reverse is actually true” (p. 157). This level of discrepancy between belief and reality points out the value of research to eliminate counterproductive standards for hiring attitudes. Rynes, Colbert, and Brown also found that the majority of respondents believed that companies are better served when they give more weight to a candidate’s values than to his intelligence, another practice that is not supported by the research (p. 162). Such subjective aspects as “values” require another method of evaluation (references, experience, or demonstrated achievement, for example). At the very least, values evaluation requires structured methodology to place it in its proper perspective and to give values their proper weight in assessing candidates for employment.

Another example provided by Rynes, Colbert, and Brown (2002) is the common misperception “that integrity tests do not work because individuals lie on them, when in reality these tests can predict job performance despite any tendencies of candidates to misrepresent themselves” (p. 168). Here again, an idea that feels correct does not hold up under statistical evaluation. Rynes, Colbert, and Brown conclude, “The power of ‘gut instinct’ and ‘chemistry’ seems to override the hard data and rational arguments” (p. 171). Administrators who rely on these instincts generally point to isolated, anecdotal examples of successful hiring to support their beliefs (Morgeson and Campion, 1997, p. 635).

To counter this personal preference in interviewing, Campion, Palmer, and Campion (1997), provide findings to support using structured interviewing practices. First, they explain that too many question types are less structured, including questions about opinions and attitudes, goals and aspirations, and self-descriptions and self-evaluations (p. 667). These questions are sufficiently ambiguous to allow candidates to present their credentials in an overly favorable manner or avoid revealing weaknesses. They focus on poorly defined traits with uncertain links to job performance (p. 668).

In an effort to counter these weaknesses, Campion, Palmer, and Campion offer the second finding: “Structure can reduce differences in perspective and might define the decision-making task such that the influence of [conflict and stress, information processing, and affect and preferences] is lessened” (p. 693). The third element of their findings concludes: “We suggest that interviews should be structured in all possible ways within any limits imposed by interviewer and candidate reactions” (p. 690), because, according to Ryan and Tippins (2004), “the research suggests that applicants will not react negatively to tools that are well developed and job-relevant” (p. 315), thereby adding to the validity and honesty of the interview.

This last facet is valuable, because “any hiring manager or researcher will agree that the number-one criterion for a useful selection device is that it provides [valid] information on who will be a good employee” (Ryan and Tippins, p. 308). It is also important to employ an instrument that has a high demonstrated validity (p. 309) so that it can provide information that is essential to a meaningful evaluation (p. 316).

A Teacher's Environment: Where Does He/She Belong?

When an administrator begins the search for a teacher, he/she must, if true success is to be achieved, consider a vast array of elements. The hiring consideration generally falls into three areas: the teacher, the school/district/community, and the specific job. Each of these components contains multiple aspects that must be considered and addressed to meet the goal: hiring the best available teacher for one particular job. When a school advertises an opening, it first offers itself for consideration. The prospective teacher examines the location, the demographics, the size of the student population, and size of the staff. The individual job comes next. The opening is for a teacher of a particular subject with particular goals (local, state, and federal). Once a teacher applies for an interview, the administrator must examine the teacher with regard to these job and district requirements. In that context come the following teacher attributes: credentials (subjects studied and grades achieved), experience and recommendations, attitude, appearance, and seemingly mandatory traits that experienced administrators have come to rely on in their hiring decisions. These last traits can become a detriment, however: "Many [administrators] look only for a certain set of traits. They have observed successful employees and they believe their traits are to be emulated by everyone. This can be a mistake" (Engel and Friedrichs, 1980, p. 87). This mistake occurs because the approach fails to look at the whole teacher with respect to the particular job. Instead, as Cable and Judge (1997) found: "Interviewer assessment of applicant fit with organizational values was related to applicant physical attractiveness, interviewer linking of the applicant GPA, and interviewer hiring recommendation" (p. 548).

A structured interview tool can assist an administrator in moving beyond personal prejudices, attitudes, and experience that may not apply to the job in question. Nicholson and McInerney (1988) discovered that “administrators were found to prize such traits as conformity, willingness to accept judgment of higher authority, attention to details, and simply being a nice person. Such qualities as academic proficiency, creativity, love of learning, patience with learners, and adaptability were sometimes de-emphasized” (p. 90). This inclusion of administrator-favored traits at the expense of student-favored traits could tend to support a compliant relationship, but not necessarily a successful, achievement-based teaching situation.

Common Problems: What Do Weak Interviews Supply?

In their research on reliability of interviewing, Engel and Friedrichs (1980), found nine common mistakes made by inexperienced or untrained (or simply unskilled) interviewers:

1. Poorly phrased questions that are not understood by the applicant
2. Purposeless questions that do not yield helpful information about the candidate (failure to have a strategy)
3. Interviewers talking too much so that the candidate is denied enough opportunity to respond to questions.
4. Interviewers who react emotionally to the candidate, thus unconsciously biasing their judgment
5. Antagonistic interviewers toward the candidate, inhibiting the candidate from adequately responding
6. Interviewers not following up on revealing candidate responses
7. Interviewer questions that go beyond the limits of proper interrogation
8. Reliance on intuition
9. Overemphasis on initial impression (p. 89)

Faced with these common problems, administrators could benefit from an interviewing tool to help them overcome these weaknesses, and they likely can't afford to spend months and years in training and research. Since these weaknesses seem focused on

preparation, the variety of applicants and jobs makes preparation complicated. As Robert Strauss (1999) explains: “Teacher selection is a complex set of procedures that includes information gathering, encoding interpretation, retrieval, and integration of information, and decision making” (p. 114). Placing these requirements in the context of traditional weaknesses of interviews yields an unpleasant image indeed.

Carrying this situation of inadequate preparation further, Terry Caldwell (1993) maintains: “Hiring qualified teachers is one of the most important steps toward providing a good education to students. Because of the large investment of time and money, teacher hiring is one of the costliest administrative duties” (p. 18). The attendant costs referred to here include the cost of paying an inappropriate teacher, the cost in time lost by all the students of that teacher, and the cost in a school’s stature, reputation, and ability to attract better teachers. As Coady (1990) explains, “Failure to hire good faculty can harm an institution for decades” (p. 6). Schools simply cannot afford this expense.

More Subtle Problems: What Throws an Interview Off Track?

Bias

Of the many subtle factors involved in interviewing challenges, bias seems most common: “To the extent that impression management [the process through which people try to control the impressions other people form of them] represents bias, it may be introducing a systematic source of inaccuracy into the interview and may negatively impact interview validity” (Posthuma, Morgeson, and Campion, p. 19). Additionally, Morgeson and Campion (1997) place the interview in human context: “As information processors, humans have limitations and biases” (p. 87). These attitudes fit both conditions of bias: a. a preference or an inclination, especially one that inhibits impartial

judgment; or b. an unfair act or policy stemming from prejudice (Ary and Jacobs 1976, p. 288). Some fairly obvious factors affect bias, as explained in the discussion by Nicholson and McInerney (1988): "...studies indicated that appearance, eye contact, gender, and attractiveness influenced the hiring decision" (p. 89). Less obvious biases, however, can be just as damaging to accuracy of an interviewer's evaluation. Of these, disability tends to impact the situation even more subtly, because interviewers may *overcompensate*: "Applicant voluntary disclosure of non-apparent disabilities and acknowledgement of apparent disability may increase ratings of employability" (Posthuma, Morgeson, and Campion, p. 74). The need to objectify, as much as possible, the direction and content of the interview becomes even more apparent when bias factors are considered.

Attitude Matching

Just as most people are interested in people like themselves, interviewers tend to give higher ratings to candidates whose attitudes mirror their own. While the desire for an attitude match is natural, and because evidence suggests that successful applicants adapt to the interviewer's communication style (Posthuma, Morgeson, and Campion, p. 67), interviewers cannot overlook applicants' skills in interviewing. Administrators perceived some candidates as more employable because they were better able to assess and respond to the impressions of administrators. Hence, they ingratiated themselves with administrators and received higher rating scores (Young, 1984, p. 47). At the same time, administrators must not *devalue* the quality of the interview experience: Teachers who interviewed well had better attendance, were more involved in voluntary in-service training, more enthusiastic, more ethnically/racially sensitive, more organized, more disciplined, had more control in the classroom, and had greater knowledge of the subject

than did teachers who interviewed less well. The interview, therefore, had a predictive quality in being able to identify applicants more likely to be more effective teachers (Loehr, 1986, p. 24). In this light, the more effective, objective, and open the interview, the more likely it is to yield meaningful data that leads to hiring an excellent teacher.

Social Factors

The subtlety of social factors comes into play because of their means of effect. Assuredly, prejudice and bias appear obvious, but research indicates that the effect comes not from the differences in such things as race, gender, intellect, and attitude, but rather in the perceived similarity between the interviewer and the applicant. In their study, Posthuma, Morgeson, and Campion (2002) explain that racial and attitudinal similarity often related to higher ratings for applicants, but that attitudinal similarity alone was generally unrelated to hiring recommendations (p. 33). Arvey and Campion (1982) suggest that interviewers higher in cognitive complexity are more likely to give higher ratings to applicants who were similar to the interviewer (p. 284). And Harris (1989) reports that interviewer-applicant sex similarity can result in interviewers asking positive questions, suggesting a confirmatory bias for applicants of the same sex and a disconfirmatory bias for applicants of the opposite sex (p. 695). Individually, these biases could negatively affect an evaluation. A candidate with a combination of these factors surely would benefit from an interview approach that reduces their subjective nature. Without an objective interviewing tool or guide, administrators are at the mercy of their own nearly undetectable biases: “Individuals act and reside in a social context and this context can influence their behavior and the processes and outcomes of the interviews” (Posthuma, Morgeson, and Campion, p. 3).

Non-Verbal Cues

Non-verbal cues require special attention. Arvey and Campion (1982) determined that nonverbal cues influence interviewer ratings, even though some evidence supports the connection between nonverbal cues and the content of the candidate's verbal information (p. 301). Another non-verbal cue comes from a judgment about personality with respect to the job available. A laboratory study demonstrated that participants gave higher ratings to applicants whose personalities fit preconceived attitudes about typical or even stereotypical personality traits found in specific occupations (Conwell, 1991, p. 28). Administrators may, therefore, tend to hire teachers who act like the typical teacher acts. Conwell (1991), however, found that even traits that were only congruent with expected traits could also influence a positive determination by the interviewer (p. 44). In this situation, also, the resulting reduction of information negatively colors the interview.

Heuristics: Expectations that Improperly Influence the Interview

In the most basic sense, interviewers bring their beliefs to the interview. Like the preconceived generalities above, specific expectations can influence the interview. These expectations manifest in an "anchoring and adjustment heuristic, which is the tendency for individuals to make judgments by selecting some standard as an anchor and then failing to sufficiently adjust from that anchor as new information is obtained" (Tversky and Kahneman, 1974, p. 1127). Such heuristic aspects surface in the primary appearance in the first interview, where "applicant personal history was more influential than applicant appearance" (Schmitt, 1976, p. 92). These determinants affect hiring, because the interview itself reveals the heuristics, while these prejudgments become a determining

factor for offering the job. As simple as this sounds, the goal is not to have a great interview; the goal is to hire a great teacher.

Lack of Structure: What Derives from Poor Interviews?

In reality, the result of the interview process is the ultimate measure of its effectiveness. Too often, unexpected influences distort the interview and lead to inappropriate hiring. Arvey and Campion (1982) summarized in *Personnel Psychology* the following twelve effects of a less-than-ideal interview:

1. General suitability ratings based on *unstructured* interviews have *low* reliability.
2. Material is not covered consistently in *unstructured* interviews.
3. Interviews are likely to weigh the same information differently.
4. Structured interviews result in higher inter-rater reliability.
5. Interview validity is low.
6. If the interviewer has valid test information available, his predictions based on the interview plus test information are usually no better and frequently less valid than the predictions based on the test alone.
7. Interviewers can reliably and validly assess intelligence but have not been shown to be effective in evaluating other traits.
8. The form of the question affects the answers given.
9. The attitude of the interviewer affects the interpretation of the interviewee's responses.
10. In *unstructured* interviews, interviewers tend to talk most.
11. Interviewers are influenced more by unfavorable than favorable information.
12. Interviewers make decisions quite early in *unstructured* interviews (p.314).

The desire for a much more structured approach to interviews is clearly merited here.

Caldwell (1993) discovered that “interviewers may typically arrive at their decisions within the first five minutes; the remainder of the interview is then used to gather supportive evidence for the initial impressions and choice” (p. 16). In this way, interviewers reduce the amount and quality of the information they gather.

The Idealistic Approach: What if the Situation were Perfect?

As Dr. Martin Luther King pointed out in his 1967 address “A Knock at Midnight”:

“Almost always, the creative, dedicated minority has made the world better” (p. 68).

Truly, some of the most exciting and productive demonstrations of good schooling come

from unique, private institutions and experimental/alternative public schools. In *If*

Teachers Were Free, Dr. Richard Renfield (1969) lists his ten critical questions for

teacher evaluation as discovered in the Potseloo, New York, schools:

1. Does the student show a growing curiosity about the world?
2. Does the student tend more often and more maturely to question his own thoughts and beliefs and what he hears and reads?
3. Does the student seek to satisfy his curiosity with growing persistence and maturity?
4. Does the student insist on verification before accepting assertions?
5. Does the student apply logic in his thinking?
6. Does the student try to be aware of his premises and evaluate them?
7. Does the student consider the likely consequences of his/her actions and beliefs?
8. Is the student becoming more creative in his thinking and his actions?
9. Is the student’s knowledge of the world becoming deeper and broader?
10. Is the student developing the strength, health, and fitness of his body and increasing his ability to do so on his own? (p. 42).

A teacher who scores well on this scale will have gone quite far toward the obvious goal of graduating a self-reliant, productive, affirmative adult. As Renfield concludes:

“Teachers at Potseloo have found that the longing to know and the longing to grow are motivations as powerful in young people as having fun, winning prestige, or passing a course” (p. 47). The trick is to find such teachers who can reach the educational interests of their students. These days, however, the vast majority of teachers must demonstrate full attention to the mandated standards of their national, state, and district superiors. The creative space is reduced more each year. Administrators, too, have to meet external standards while serving their individual populations, and since teachers work most with

their principals (second only to their students), it makes sense that the authority and *responsibility* for hiring and keeping great teachers should fall to the person closest to the action. Seeking to validate the employment interview, Posthuma, Morgeson, and Campion (2002) conclude that “although there are some differences between interviewers, it appears that differences can be mitigated by increasing interview structure or by increasing interviewer accountability” (p. 75). By adding a formal interviewing tool to the accountability, school districts can, within a single term, spot strengths and weakness in their hiring procedure and make corrections.

The next item of impact is court decisions and legislative actions. Posthuma (2002) found that these legal factors and districts’ compliance with them might very likely affect interviewers and organizations. Schmitt (1976) suggests that “pressure for quotas would affect interview outcomes and that it was becoming increasingly important for interviews to be validated as job-related selection tools” (p. 39). Again, a formal interview tool can help insure that information in support of the court and legislative requirements is not overlooked.

Conclusion: Using the Right Tools for the Job

Huffcutt et al. (1996) found that interviewer ratings were better at reflecting cognitive ability with situational interviews than behavioral description interviews when the interview had a high level of structure (p. 462). Barrick, Patton, and Haugland (2000) added: “It is now widely accepted that the predictive validity of an interview is increased when the interview questions and scoring procedures are structured and the content is job related” (p. 929). And although administrators want to know the beliefs and personalities of candidates (Metzger and Wu, 2003), an interview is not really able to “measure” these

qualities (Barrick, Patton, and Haugland, p. 943) and renders measures of such factors as personality and preference invalid (p. 945). The structured interview does a good job, however, of measuring job fit and fitness (p. 948).

The clear consensus of the literature is that unstructured interviews are filled with potential pitfalls for the administrator. The increasing variety of demands on modern schools and teachers leaves little room for casual/informal efforts in hiring. With the current availability of resources for structured interviews and the increasing supporting research in favor of structured interviews, administrators have significant amounts of information for making a useful decision concerning interview instruments.

III. Method

The current study compared results of teacher interviews completed within a month period of time using the ICIS instrument with student evaluations of these same teachers collected at the end of the year. The goal of this study was to examine the *concurrent* validity of the ICIS using student satisfaction as the dependent measure. The results of the student survey were correlated with scores obtained from the ICIS teacher interviews.

The school chosen for this study provided a richly diverse research sample of students and teachers. Located in an area including urban core, industry, suburban developments, and traditional neighborhoods, the high school, at the time of the study, served a variety of social, ethnic, economic, and academic groups. The student population at the time of the study was 48% male and 52% female; 60% White, 19% African-American, 12% Latino, 8% Asian, and 1% American Indian. The free/reduced lunch population was high (44.62%); the mobility ratio was high (20.3%); the school and the district were old (94 years); and the school was relatively large (1352 students in four grades). The grade point averages of students selected to participate in this study (on a 4-point scale) ranged from 0.01 to 4.0. Eighty percent of the research sample were freshmen, and twenty percent were sophomores.

The teachers represented a broad variety of subjects and disciplines, experience, and ethnicity. They all taught required classes in academic subjects (English, math, science, and social studies). The teachers surveyed for this study were 38% male and 62% female. Their average experience was 8.2 years and ranged from one to twenty-eight years. They were 78% White, 15% African-American, and 7% Hispanic. The school employed 127 secondary teachers in a district of 506 secondary teachers.

The Student Survey Instrument

The student survey was taken from the published article, *The Predictive Validity of Student Evaluations in the Identification of Meritorious Teachers* (Ebmeier, Crawford, and Jenkins, 1991), a study of students' ability to acknowledge strengths of classroom teachers. School district personnel were looking for a reliable, efficient, cost-effective model to help determine which teachers should receive a performance bonus. Using nine criteria [see Appendix A], a team of education professionals developed an observation instrument to be used by external evaluators (professionals from outside the district). This team also created a survey instrument to be completed by students. Its thirty items form the "Steps to Excellence Student Questionnaire" [see Appendix B]. This instrument was based on the same nine criteria as the teachers' evaluation instrument. The student survey showed a significant positive correlation with evaluations conducted by external experts. This student survey has been shown to be particularly useful, because it is rooted in measures that students can both appreciate and fairly judge. Additionally, the student survey evaluates teacher behaviors that clearly lead to student satisfaction with their teachers [Appendix B]. The survey committee that designed this instrument included eight veteran teachers with master's degrees, five administrators, and a school board member.

Ebmeier, Jenkins, and Crawford (1991) explain that the construct validity was examined through a factor analysis that shows the questionnaire to be unidimensional. Reliability of the student instrument is high (Cronback's $\alpha=0.93$), because it is based on the commonality of questions and the high correlation of the questions. A scale score

for each respondent is derived by summing the answers across all 30 individual questions (Ebmeier, Jenkins, and Crawford, p. 351).

The ICIS Interview Instrument

The interview questions for the ICIS are divided into four major categories: Working with Others, Knowledge of Content, Knowledge of Teaching, and Knowledge of Students. The Knowledge of Teaching scale is further subdivided into Delivery of Instruction, Planning, Climate Development, Assessment, and Interactions.

Question selection and development for the ICIS instrument came from two documents created by national studies. A national commission of school personnel officers provided material for *Teachers of the Future* (1997), the first publication informing the instrument's content. Following a two-year review of existing literature and advice from practitioners, the commissioners selected nine knowledge areas and eleven skill areas necessary for all teachers (p. 47).

The second major document used as the basis for constructing the interview questions, *Praxis III: Classroom Performance Assessments* (Dwyer, 1998), was developed for use with beginning teachers. Created for a variety of evaluation purposes (licensure and professional development), *Praxis* builds on a base using knowledge of general principles, particular students and their backgrounds, content and its organization, and other specific knowledge and skills necessary for effective teaching. The Educational Testing Service, in cooperation with a group of practicing teachers and under the direction of a National Advisory Committee, carried out research and development activities over a ten-year period. Creation of rubrics to score the interviews followed the question construction. Using descriptions of effective/ineffective practices that came

from the process-product research conducted over thirty years, example responses [see Appendix C] were developed to guide the interviewer's evaluation (Ebmeier, p. 6). Equipped with the questions and the rubrics, the interviewer can ask questions, score the responses, and then refer to the computer's record keeping and reporting to determine which areas require further questioning and exploration.

Content validity was supported by adherence to constructs important to both of the foundational documents, thus assuring compliance with two national commissions. Internal reliability of all subscales was over 0.70 with the total score in excess of 0.90. Importantly, the interviewer of the teachers completed a training program designed for the ICIS series of instruments and a demonstrated over 90% mastery of the instrument and method of interviewing prior to conducting the interviews.

Structure of the Data Matrix

After students completed their surveys, their data were joined by objective data gathered from their class records. Teachers furnished specific data for their students so that the individual characteristics could be correlated with the teachers' ICIS scores. Student names, identification numbers, or any other information that would connect a student with his/her survey were kept separate from the data to maintain student anonymity. The "teacher" portion of the tracking device included an identification code, the teacher's age, and the subject taught. The "student" portion included gender and average overall GPA. The unit of analysis was the individual student.

Data Collection Methods

The forty teachers were selected according to the premise that students in required classes would provide a more accurate representative sample of abilities and attitudes

than students in elective classes. The subjects chosen were math, science, social studies and communication arts. The forty teachers were selected at random from among all the core teachers, and all those chosen volunteered to participate. All teachers were interviewed using the ICIS instrument in an authentic setting by the study's author (either in person or by phone) between May 10 and June 20, 2009. All students in each class completed the surveys in order to provide the best possible sample (N=1039). The author of this study administered all the surveys to all the classes. A time frame for completing the survey was established in the spring to give the students the greatest possible experience with the teachers before they completed their surveys. All surveys were completed within the one-week period between April 13 and 17, 2009.

Data Analysis

This study compared the individual student survey average scores with the ICIS ratings of the teacher for which he or she completed the survey. Each student score was matched with his/her teacher's total ICIS rating and the teacher's rating for each of the four subgroups. The study also correlated each of the teachers' individual, personal factors (age, experience, and gender) with the student's survey score and teacher's gender and experience level.

IV. Results

Descriptive Statistics

The following table demonstrates the results of the overall examination of data. The numbers relating to the ICIS and the student survey offer the most relevant information and represent the greatest consistency (with a .85 difference between the highest and lowest scores in standard deviation among the ICIS respondents on the total score).

Table 4-1
Descriptive Statistics of the ICIS and Student Survey

	Mean	Standard Deviation	Variance	Range	Minimum	Maximum
Teacher Age	36.9153	11.06342	122.399	35	24	59
Teacher Experience	8.2926	6358083	43.307	27	1	28
Student Gender	1.4601	0.49864	0.249	1	1	2
Teacher Gender	1.5871	0.49259	0.243	1	1	2
Student GPA	2.2555	0.8988	0.808	3.97	0.01	3.98
ICIS W/O	2.5392	0.29068	0.084	1	2	3
ICIS K/C	2.5132	0.23639	0.056	0.78	2.1	2.88
ICIS K/S	2.5132	0.23639	0.056	0.78	2.1	2.88
ICIS K/T	2.3366	0.33246	0.111	1.25	1.75	3
ICIS Total	2.47724	0.22503	0.051	0.85	2	2.85
Survey Average	2.5157	0.69186	0.479	3	1	4

Concurrent Validity

The indicators of validity of the ICIS rest predominantly on the Pearson correlation coefficient. This measurement procedure provides a single number that describes the connection between two sets of variables: in this case, the teacher's ICIS score and the student-survey average score. The Pearson correlation is especially useful because it can effectively blend both positive and negative results, simplifying the examination of those results. The Pearson correlation was calculated through use of the Statistical Package for the Social Sciences (SPSS), a computer program that permits many types of statistical analysis.

As can be observed from Table 4-2, the Pearson correlation between the total (overall) ICIS score and the overall student survey average score (.268) demonstrates a significant positive relationship, since the two-tailed correlation is significant at $p < .01$. This result comes from over 1,000 surveys of thirty questions each and an average of 30 questions from each of the 40 teacher surveys for a combined total of over 31,000 items. The resulting score, as defined by students' attitudes about their teachers, is an indicator of the ICIS's effectiveness in identifying teachers' ability to teach in a way that will produce student satisfaction and, by extension, good student performance.

In fact, evaluating the four segments of the ICIS yields similar Pearson correlative results concerning the relationship between the administrator's judgment and the students' judgment: Knowledge of Content ($p < .177$); Knowledge of Student ($p < .201$); Knowledge of Teaching ($p < .265$); and Working with Others ($p < .197$). The highest score (Knowledge of Teaching) seems encouraging for two reasons: first, the sub-groupings of Knowledge of teaching provide the greatest variety of questions; and, second, students demonstrated the ability to recognize their teachers' ability to teach.

TABLE 4-2
Pearson Correlation Coefficients: ICIS Ratings and Student Survey Averages

	Teacher Age	Teacher Experience	Teacher Gender	Student Gender	Student GPA	Student Survey AVG
Teacher experience						
Pearson Correlation	.644**					
Sig. (2-tailed)	.000					
N	1039					
Teacher gender						
Pearson Correlation	-.046	.184				
Sig. (2-tailed)	.137	.000				
N	1039	1039				
Student gender						
Pearson Correlation	.081**					
Sig. (2-tailed)	.009					
N	1039					
Student GPA						
Pearson Correlation	.070*			210**		
Sig. (2-tailed)	.024			.000		
N	1039			1039		
ICIS Working with others						
Pearson Correlation	-.005	.202**	.155**			.197**
Sig. (2-tailed)	.866	.000	.000			.000
N	1039	1039	1039			1039
ICIS Knowledge of Content						
Pearson Correlation	.092**	.358**	.355**			.177**
Sig. (2-tailed)	.003	.000	.000			.000
N	1039	1039	1039			1039
ICIS Knowledge of Student						
Pearson Correlation	.092**	.358**	.134**			.201**
Sig. (2-tailed)	.003	.000	.000			.000
N	1039	1039	1039			1039
ICIS Knowledge of Teaching						
Pearson Correlation	-.199**	.084**	.313**			.265**
Sig. (2-tailed)	.000	.006	.000			.000
N	1039	1039	1039			1039
ICIS Total Score						
Pearson Correlation	-.070*	.225**	.294**			.268**
Sig. (2-tailed)	.024	.000	.000			.000
N	1039	1039	1039			1039
Student Survey Average						
Pearson Correlation	-.084**	.072*	.070*	.015	.077*	
Sig. (2-tailed)	.007	.021	.023	.632	.013	
N	1039	1039	1039	1039	1039	

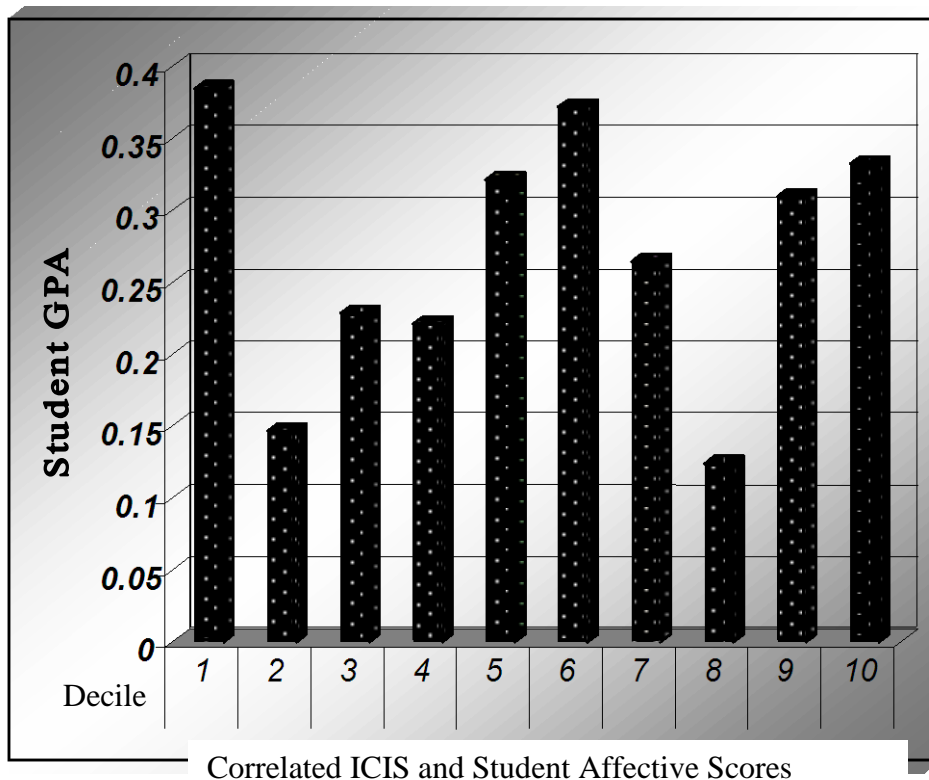
**Correlation is significant at the 0.01 level (2-tailed)

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

Examination of Aspects of Bias

To explore the possibility that lower GPA students have less accurate estimates of teaching competence (assuming the ICIS scores were accurate), the students were divided into ten groups (deciles) based on their GPA and then each group was used as a sample for ten separate calculations of the correlation of the ICIS ratings of their teachers with the student evaluation of the teachers. Chart 4-3 presents the results.

Chart 4-3
 Correlations of the ICIS and Student Satisfaction Scores
 Grouped by Student GPA Deciles



Decile 1 represents the lowest 10% of GPAs. Students were first grouped by GPA and then the correlation between the ICIS scores and Student Affective Scores were calculated within each of the deciles.

While the above table shows a slight skew toward the higher GPA deciles, the overall view remains random. In other words, the students' GPA appears to have little

relationship to the Pearson coefficient between the student survey and the ICIS total score. The absence of relationship, while in contrast to an expectation that higher-GPA students would rate their teachers with more similarity to administrators than lower-GPA students, was interesting with the explanation remaining a mystery and subject for additional study.

V. Discussion

The ICIS program seems to help identify teachers whom school districts can reasonably and confidently expect to build positive relations with their students. Obviously, however, no single selection device can predict outcomes with perfect accuracy. Any interview can only be one factor in the overall evaluation of teacher candidates; the process must also include previous work history, references, transcripts, and observations. The ICIS seeks to objectify as much as possible a highly subjective event (the interview) and serves to guide administrators by providing just such a context, and the data in the current study support the value of its application.

Just as the hiring process demands evaluation of multiple factors, the ICIS itself requires evaluation from multiple directions. The current study, therefore, brings together important characteristics of the teachers, their scores on the ICIS interview, and student affective responses based on a full year of contact with the teacher. Unlike observations, which occur as a single incident (or, at best, two or three incidents), the student evaluators in this study bring the invaluable perspective of time, with its variety of experiences and conditions, to the equation. The student survey's significant positive correlation thus carries great weight in support of the ICIS's effectiveness. Also, the anonymous student responses (1039) are numerous and unaffected by the teachers' manipulation, while references are, by design, few and carefully selected by the candidate. Here again, the student survey, while not necessarily "expert" in qualification, still offers a breadth and candor often missing in references chosen by the candidate. Furthermore, the 1991 study of the *Predictive Validity of Student Evaluations* cites various other studies that find student evaluations "stable over time," "reliable," and as

evidencing “good discriminant and convergent validity” (p. 343). The Ebmeier, Jenkins, and Crawford study itself found that “students can successfully discriminate between the meritorious and nonmeritorious teachers” (p. 353).

Other studies have taken a different slant on validating the ICIS. Reik’s 2007 research studied the connection between the Missouri Assessment Project (MAP) scores and the ICIS. Reik found a correlation of .454 between the MAP scores and the ICIS weighted Total Score by comparing the three-year average scores for elementary age students and results of the ICIS interviews. The lowest correlation of his findings (Knowledge of Content) was a significant .293. The student opinion ratings of the current study add to Reik’s findings, offering one more facet of validation for the ICIS.

In similar studies, Green (2005) and Dillon (2006) showed the validity of standardized questions in evaluating candidates. Green combined information from individual school counselors, counseling associations, and university programs. Dillon created coordinated questions for interviewing paraprofessionals. Both these studies help to validate the ICIS concept of a formalized interview instrument.

A study by Gardner (2009) demonstrated a higher correlation than that of the current study between the student survey and the ICIS interview scores: Working with Others (.494), Knowledge of Content (.539), Knowledge of Teaching (.532), Knowledge of Students (.617), and ICIS Total (.635). The primary and significant difference between the student sample in Gardner’s study and that of the school in the present study may be traced through the Free/Reduced Lunch population. The Gardner school’s F/R percentage was 15%, while the present study’s school’s F/R percentage was 44.62%. The lower socio-economic condition of the students of the school in the present study generally

results in a higher mobility ratio and lower attendance average and lower grade point average. These three factors (F/R, mobility, and GPA) may well lead to lower interest in school in general and lower opinion of teachers in particular, resulting in a lower correlation between administrator perceptions (ICIS interview score) and student attitudes (student survey score). Two other possibilities exist that might explain the higher student survey scores for the Gardner teachers: higher-achieving students generally take more care in completing the surveys, and the Gardner students were older than the students of the current study. The current study offers results that add to the research picture by examining the attitudes of an unstructured sample of students in the most authentic setting possible.

Limitations of the Current Study

Convenience Sample

The idea of a convenience sample certainly applies to a study conducted within the confines of a single school (in this case an urban high school of 1532 students and 127 teachers). This might make this sample quite different from another sample. Further, in an “authentic” setting, “interviews would be administered to pre-service teachers which are not subjected to . . . enhancing [experiential] elements.” Acknowledging the fundamental truth of these statements, studies still show that the “diversity of challenges present in an urban high school provides researchers with a quite practical ‘laboratory’ for empirical study” (Cohen, p. 12). Additionally, the broader mix of academically (for a variety of reasons) challenged students offers a better testing ground for a teacher’s ability.

Finally, although the convenience of interviewing and evaluating experienced teachers may fall outside the authentic setting, any realistic interview process would include teachers with widely varying degrees of experience. The school of the current sample, 94 years old, has teachers with experience ranging from zero to twenty-eight years. That range and variety of experience helps support the validity of the current study even though it was conducted within the framework of a convenience sample.

Validity

Some researchers construct studies that rely as much as possible on “raw material,” asserting that the influence of experience colors the sample; they prefer to interview recent graduates and then follow their experience after the predictions have been made about their potential. This method of study, however, eliminates experienced teachers from the examination of the interview instrument, and administrators must interview many experienced teachers. The evaluation of a study’s validity, either predictive (as some researchers favor) or concurrent (as in this study), therefore, depends more on the effective interpretation of the study rather than on the nomenclature (Cohen, p. 19). Ebmeier (2006) explains that “time commitment issues associated with multi-year studies” have restricted the possibility of timely examination of the current interview instrument to concurrent validity. Indeed, given the newness of the current interview instrument, establishing concurrent validity may be the most practical way to add credence to the ICIS as an interview tool.

Suggestions for Further Research and Improvements

Institutional Impact

In today's numbers-driven educational environment, even successful, experienced teachers feel the pressure to perform according to a state-sponsored standard. Recent research suggests teachers strive to answer interview questions in a way that shows their understanding of state-required testing standards (Reik, 2007, p. 42). Typically, those answers score only in the median range on the ICIS, thus potentially lowering the scores of teachers interviewed in districts that use state testing as curriculum guidelines. Even though the current study does not deal directly with state testing, the fact that state testing affects so much of current curriculum decision making cannot be dismissed as a factor in teachers' performance during an interview. Comparisons between interviews conducted for public-school and private-school jobs could be useful in countering this factor, because it would provide results free of state-testing bias.

Administrator's Views in Relation to Student Views

The current study compared scores from an employee interview with the students' affective evaluations of their teachers. While a student's response to the teacher is by far the most critical aspect of teacher suitability, one could make a case for the importance of a teacher's suitability for the organization/school and its personality. Cable and Judge (1997) found that interviewer assessment of an applicant's fit with an organization's values has not been adequately studied as a valid predictor of future job performance. The present study answers that question to some degree by examining teachers of varying lengths of service, but the more specific question of "institutional fitness" would require a

different focus of examination, one that included attitudes of parents, other teachers, and administrators in addition to the responses of the students.

Academic Preparation

A simple addition to the present study would be to create an objective descriptor for a teacher's academic preparedness. Factors like degrees earned, hours of credit in the teacher's subject assignment, schools of education or professional training (like Teach for America programs), or any similar comparisons could be a part of the initial teacher survey. The true goal, however, would be to show how these areas of concern, while probably meaningful for interviewing and hiring, might not impact the students' survey results. Nevertheless, examining the connection between a teacher's preparedness and student impressions of him/her could be useful.

Conclusion

The Interactive Computer Interview System has continued to earn support from increasing numbers of studies. The connection to previously validated student surveys (Ebmeier, Jenkins, and Crawford, 1991), the Missouri Assessment Project (Reik, 2007), and various counseling and school psychology departments (Smith, 2005), has shown correlative effectiveness in various and diverse applications. Even given the cautionary note from the developer of the ICIS, "Despite its wide acceptance, the employee interview has definite limitations. Researchers over the last fifty years have cautioned employers not to overemphasize the interview's importance. Estimates indicate that even the best interview process generally can only account for about fifty percent of the variance in predicting future employee quality" (Ebmeier 2003, 22), the ICIS has moved far in the direction of ensuring that the "fifty percent" has the greatest possible relevance

in the highest number of situations. A large school district with many interviews to conduct on various days, in various locations, by various administrators, could gain greatly by using the objectified and consistent results of a computerized interview tool. Giving proper weight to the other factors in hiring (grades, academic depth, experience, past performance, and references), the computer-assisted ICIS interview could be an important factor in allowing administrators to make decisions with more confidence when hiring new teachers.

VI. References

- Ary, D. & Jacobs, L. C. (1976). *Introduction to Statistics: Purposes and Procedures*. New York: Holt, Rinehart and Winston.
- Arvey, R. D. (1995, May). Panelist. In Campion, MA, and Palmer, DK (Chairs), *Taking stock of structure in the employment interview*. Panel discussion presented at the Tenth Annual conference of the Society for Industrial and Organizational Psychology, Inc. Orlando.
- Arvey R. D., & Campion J. E. (1982). The employment interview: A summary and review of recent research. *Personnel Psychology*, 35, 281-322.
- Barrick, M. R., Patton, G. K., & Haugland, S. N. (2000). Accuracy of interviewer judgments of job applicant personality traits. *Personnel Psychology*, 53: 925—951.
- Cable, D. M., & Judge, T. A. (1997). Interviewers' perceptions of person-organization fit and organizational selection decisions. *Journal of Applied Psychology*, 82: 546—562.
- Caldwell, T. G. (1993). Hiring excellent teachers: current interviewing theories, techniques, and practices. Unpublished thesis. California State University, Long Beach.
- Campion, M. A., Campion, J. E., & Palmer, D. K. (1997). A review of structure in the selection interview. *Personnel Psychology*, 50: 655—702.
- Coady, S. How to Hire Good Faculty. (May 20-23, 1990). Paper presented at the annual International conference on Teaching Excellence and conference of Administrators.
- Conwell S. L. (1991). Inferential judgment in the employment interview. (Doctoral dissertation, Texas Tech University, 1990). *Dissertation Abstracts International* 52: 1428.
- Cohen J. (1977). *Statistical power for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.
- Dillon, A. (2006). The development of an interview system to identify quality paraprofessionals. Unpublished doctoral dissertation, University of Kansas.

- Dwyer, D. (1998). Psychometrics of Praxis III: Classroom performance assessments, *Journal of Personnel Evaluation in Education*, 12 (2), 163-187.
- Ebmeier, H. (2003) *Interactive Computer Interview System Technical Manual*, American Association of School Personnel Administrators, Version 1.
- Ebmeier, H., Jenkins, R., & Crawford, G. (1991). The predictive validity of student evaluations in the identification of meritorious teachers, 341—357. *Journal of Personnel Evaluation in Education*, 4 (4), 341-357.
- Emly, K. (1994). The effect of interview format on principal opinion of teachers. Unpublished doctoral dissertation, University of Kansas.
- Engel, R. A., & Friedrichs, D. (1980) The interview can be a reliable process. *NASSP Bulletin* 64 (432), 85—91.
- Fenstermacher, G. D., & Richardson, V. (2005). On making determinations of quality in teaching. *Teachers College Record* 107(1), 186—212.
- Gardner, R. C. (2009) A study of the construct validity of the Interactive Computer Interview System (ICIS) using student evaluations as the outcome measure. Unpublished doctoral dissertation, University of Kansas.
- Gollub-Williamson, L. G., Campion, J. E., Malos, S. B., Roehling, M. V., Campion, M. A. (1997). Employment interview on trial: Linking interview structure with litigation outcomes. *Journal of Applied Psychology* 82: 900—912.
- Green, M. (2005). Interviewing standards for hiring school counselors. Unpublished doctoral dissertation, University of Kansas.
- Harris, M. M. (1989). Reconsidering the employment interview: A review of recent literature and suggestions for future research. I: 691-726.

- Harris, N. M., McIntire, K. E., Littleton, V. C., Jr., & Long, D.F. (1979). *Personnel administration in education*. Boston: Allyn and Bacon.
- Huffcutt, A. I., Roth, P. L., & McDaniel, M. A. (1996). A meta-analytic investigation of cognitive ability in employment interview evaluations: Moderating characteristics and implications for incremental validity. *Journal of Applied Psychology, 81*: 459—473.
- King, M. L., Jr. A knock at midnight in (1998) P. Holloran & C. Carson (Eds.), *A Knock at Midnight* (pp. 61—71). New York: Intellectual Properties Management in Association with Wonder Books.
- Loehr, P. (1986). Relationships between teacher interview scores and on-the-job performance.” Paper presented at the Annual Meeting of American Association of School Personnel Administrators.
- Manning, .H., & Payne, B. D. (1984). Student teacher personality as a variable in teacher education. *The Teacher Educator, 20* (2): 2—12.
- McWalters, P. (2004). Increasing achievement for all children.” *The New England Reading Association Journal 40* (2): 68—70.
- Metzger, S. A., & Wu, M-J. (2003, April 21—25). Commercial teacher interviews and their problematic role as a teacher qualification. Paper presented at the American Educational Research Association Annual Conference, Chicago.
- Morgeson F. P., & Campion, M. A. (1997). Social and cognitive sources of potential inaccuracy in job analysis. *Journal of Applied Psychology, 82*: 627-655.
- Nicholson, E. W., & McInerney, W. D. (1988). Hiring the right teacher: A method for selection. *NASSP Bulletin 72*(511): 88—90.

Overview of *A Guide to Education and No Child Left Behind* October 2004. US Department of Education.

Posthuma, R. A., Morgeson, F. P., & Campion, M. A. Beyond employment interview validity: A comprehensive narrative review of recent research and trends over time. *Personnel Psychology*, Durham: Spring 2002, Vol. 55, Iss. 1; pg. 1, 81 pgs.

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 (MAP) as the outcome measure. Unpublished dissertation, University of Kansas.

Renfield, R. (1969). *If Teachers were Free*. Washington, D.C.: Acropolis Books.

Ryan, A. M., & Tipipins, N. T. (2004). Attracting and selecting: what psychological research tells us. *Human Resource Management*, 43 (4): 305—318.

Ryan, P. M., & Alcock, M. A. (2002). Personal and interpersonal attributes in selecting teachers. *Action in Teacher Education* 24 (1), 58—67.

Rynes, S. L., Colbert, A. E., & Brown, K. G. (2002). HR professionals' beliefs about effective human resource practices: Correspondence between research and practice. *Human Resource Management*, 41: 149—174.

Schlechy, P. (1981). Do academically able teachers leave education?: The North Carolina case. *Phi Delta Kappan* 63 (2): 106—112.

Schmitt, N. (1976). Social and situational determinants of interview decisions: Implications for the employment interview. *Personnel Psychology*, 29: 79-101.

Slosson, J. (1999). Hiring the right people. *High School Magazine* 7(2): 27—30.

- Smith, V. W. (2005) The development and field testing of a school psychologist interview instrument. Unpublished doctoral dissertation, University of Kansas.
- Smith, T. M., Desimone, L. M., & Ueno, K. (2005). "Highly qualified" to do what? The relationship between NCLB Teacher Quality Mandates and the state of reform-oriented instruction in middle school mathematics." *Educational Evaluation and Policy Analysis* 27 (1): 75—109.
- Strauss, R. P. (1999). "Who gets hired to teach? The case of Pennsylvania." *Better Teachers, Better Schools*, New York, The Thomas B. Fordham Foundation: 103—116.
- Teachers of the Future: A Continuous Cycle of Improvement*. (1997). Olathe, Kansas: American Association of School Personnel Administrators.
- Tengler, C. D., & Jablin, F. (1983). Effects of question type, orientation, and sequencing in the employment screening interview. *Communication Monographs*, 50: 59—62.
- Tversky A, & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, 85: 1124-1131.
- Young, I. P. (1984). The effects of interpersonal performance style in simulated teacher selection interviews. *Journal of Research and Development in Education* 17(4): 43—51.

VII. Appendices

Appendix A: Criteria for Data Collection by External Team of Experts

1. The teacher plans and organizes instruction.
2. The teacher uses a variety of instructional techniques, methods, and media appropriate to the class and subject matter.
3. The teacher communicates effectively and positively with students.
4. The teacher uses effective instructional deliver techniques.
5. The teacher monitors and evaluates student learning.
6. The teacher creates and maintains a positive learning environment.
7. The students participate in class activities.
8. The students seek additional information.
9. The students take responsibility for learning.

Appendix B: Student Survey Questions

Steps to Excellence Student Questionnaire

1. My teacher shows enthusiasm about the subject he/she is teaching.
2. My teacher keeps me interested in my school work.
3. My teacher explains to us what we are supposed to learn.
4. My teacher gives clear directions for doing our school work.
5. My teacher explains things again if I don't understand.
6. My teacher gives me a chance to say what I think in class.
7. My teacher tells me why the things we learn in class are important.
8. My teacher keeps me interested and working during class.
9. My teacher notices if I am not paying attention.
10. My teacher uses more than one way to teach us things.
11. My teacher gives us enough practice before we move on.
12. My teacher helps me to learn how I can do a better job on my school work.
13. My teacher encourages me to do the very best I can on my school work.
14. My teacher treats me friendly.
15. My teacher grades my work carefully and fairly.
16. My teacher returns my work quickly.
17. My teacher works with students the whole class period.
18. My teacher lets me know if my behavior is right or wrong.
19. My teacher keeps students well-behaved.
20. When students misbehave, my teacher disciplines them in a fair way.
21. My teacher lets me know when I do well on my school work.
22. My teacher makes me feel good about my work.
23. My teacher presents new material in ways that I can understand.
24. My teacher says positive things to students.
25. My teacher shows that he/she cares about me.
26. I enjoy learning in this class.
27. My teacher expects us to work on our school work.
28. My teacher has materials well-organized.
29. My teacher holds students responsible for their school work.
30. I feel comfortable asking my teacher questions (pp. 348—350).

Appendix C: ICIS Scoring Rubrics

Cluster	Levels of Development
Working With Others	<p><u>Level 1:</u> 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.</p> <p><u>Level 2:</u> Focuses on own classroom but sees the importance of school coordination and interactions with others for the “good of the school”.</p> <p><u>Level 3:</u> Altruistic motivation is the driving force for these teachers. Concerned with the larger good. Great respect for “social capital” idea.</p>
Knowledge of Content	<p><u>Level 1:</u> Minimal knowledge. Lacking basic college coursework in much of the field</p> <p><u>Level 2:</u> Adequate knowledge base typical of a 36 hour college major in the subject field.</p> <p><u>Level 3:</u> Expert knowledge typical of a major with more than 50 hours in the field.</p>
Knowledge of Instruction (Delivery of Instruction)	<p><u>Level 1:</u> Coherence of an instructional delivery plan is typically lacking.</p> <p><u>Level 2:</u> Teachers consider multi-part segments of the lesson, how they fit together, why the sequence is important.</p> <p><u>Level 3:</u> Teacher considers multi-part behaviors that are selected and executed based on on going analysis of classroom events.</p>
Knowledge of Instruction (Planning)	<p><u>Level 1:</u> 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.</p> <p><u>Level 2:</u> The teacher begins to think about what he or she wants the students to be able to do at defined points in the lesson.</p> <p><u>Level 3:</u> Teacher incorporates branching designs into the planning of their lessons such that they could easily vary the content and method based on classroom feedback.</p>
Knowledge of Instruction (Climate Development)	<p><u>Level 1:</u> General lack of student attentiveness to academic tasks.</p> <p><u>Level 2:</u> Students are attentive to teacher directed instruction but not necessarily when working in unsupervised groups.</p> <p><u>Level 3:</u> Students self-regulate behavior commensurate with the learning goals.</p>
Knowledge of Instruction (Assessment)	<p><u>Level 1:</u> Assessments viewed as a means of student control. Assessment are infrequent, rarely measure important concepts, poorly designed, with results rarely affecting instruction.</p> <p><u>Level 2:</u> Assessments are viewed as a means of grading and to some extend providing feedback to the teacher about instruction.</p> <p><u>Level 3:</u> Assessments are viewed as a means of diagnosing individual student process and product understanding.</p>

(table continues)

Appendix C: ICIS Scoring Rubrics

Cluster	Levels of Development
<p>Knowledge of Teaching (Interactions)</p>	<p><u>Level 1</u>: Information and interactions are often confusing for students. They have difficulty knowing what is expected.</p> <p><u>Level 2</u>: 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.</p> <p><u>Level 3</u>: Teacher presents information in a way that increases the chances students will comprehend. Teacher thematically connects statements and links student responses to prior material.</p>
<p>Knowledge of Students</p>	<p><u>Level 1</u>: Minimal teacher knowledge of educational psychology. Teacher lacking basic exposure to students of this age or background.</p> <p><u>Level 2</u>: Academic knowledge, student teaching experience, and non-school related teaching exposure to students such as summer camps or church school.</p> <p><u>Level 3</u>: Academic knowledge, teaching experiences in the same context, community experience, and out-of-class contact in students' environment.</p>

Appendix D: Teacher Scores on the ICIS

Number	Age	Experience	Gender	W/O	K/C	K/S	K/T	Total
1	32	7	M	2.25	2.25	2	2.13	2.15
2	30	5	F	2	2.6	3	2.88	2.34
3	27	3	F	2	2.25	3	2.63	2.45
4	37	12	F	2.4	2.75	3	2.63	2.67
5	52	28	F	1.6	2	1.75	1.75	1.76
6	51	27	F	1.4	1.75	1.6	1.63	1.59
7	28	4	M	1.4	2.25	1.75	1.63	1.71
8	31	6	F	2	2.75	2.6	2.5	1.59
9	32	9	F	2	1.72	2.6	2.6	1.71
10	29	5	F	2.25	3	2.75	2.693	2.48
11	27	2	F	2.6	2.75	2.6	2.63	2.13
12	39	4	M	2.75	3	2.4	2.38	2.65
13	47	21	M	2.25	2.4	2.25	2.38	2.64
14	38	16	F	2.6	3	2.25	2.5	2.57
15	37	13	F	2.25	2.6	3	2.63	2.62
16	59	26	F	2.25	2.6	2.75	2.13	2.38
17	44	18	M	2.6	2.75	2.25	2.5	2.52
18	27	3	M	2.25	2.75	3	2.88	2.75
19	32	10	F	2	2.4	2.25	2.38	2.29
20	45	19	F	2.25	2.4	2.75	2.5	2.48
21	40	18	F	2	2.75	2.75	2.25	2.4
22	45	23	F	1.75	2.25	2	1.55	1.95
23	55	26	F	2	2.25	2	1.88	2
24	35	12	F	2.25	2.75	2.75	2.75	2.65
25	31	8	F	2.25	2.6	2.75	2.63	2.57
26	32	12	M	1.25	1.75	2	1.88	1.75
27	34	3	M	1.83	2.25	1.75	2	1.95
28	39	17	M	2	2.75	2	2	2.15
29	36	14	M	1.83	2.4	2.25	2.38	2.22
30	50	28	F	2.25	2.75	2.6	2.75	2.62
31	31	9	M	1.75	2.75	2.75	2.25	2.35
32	41	19	F	2.25	2.75	2.25	2.38	2.4
33	39	17	M	1.75	2	1.75	1.88	1.85
34	49	26	M	1.4	2.25	1.75	1.63	1.59
35	42	18	F	2	2	2	2.25	2.15
36	40	17	F	1.25	1.75	2	1.75	1.7
37	32	9	M	2	2.75	2.25	2.25	2.27
38	47	24	F	2.25	2.75	3	2.88	2.75
39	33	5	F	2	2.6	2.6	2.75	2.55
40	53	28	M	1.85	1.75	2	1.88	1.75