

CHAPTER 3. METHODS

Chapter 3 will describe the research methodology used to develop and field test a survey designed to measure education technology leadership. This chapter provides detailed information on the methodological approach of this research. The study's research design is presented, including the study's phases and the associated participants, the procedures used for data collection, and the statistical methods used to analyze the data. A research design issues section and a section discussing human subject research in relationship to this study are also included in this chapter.

Methodological Approach

This study used quantitative research methods. Quantitative research has been defined as “empirical research in which the researcher explores relationships using numeric data” (Fraenkel & Wallen, 1993, p. 5). Quantitative research methods are the process of developing systematized knowledge gained from observations that are formulated to support insights and generalizations about the phenomena under study (Lauer & Asher, 1988). Quantitative methods are an approach to research in which the investigator primarily uses post-positivist claims for developing knowledge (Creswell, 2003; Crotty, 1998).

Post-positivism followed the traditional positivist theoretical approach, but accounted for the fact that one may never know absolute truth in seeking knowledge. There are inherent assumptions in the post-positivist perspective (Phillips & Burbules, 2000), which include:

- Knowledge is seen as conjectural and supported by the strongest warrants of the Time, but can be subject to reconsideration at any time.

- Warrants for claims are made and examined; if the warrant is strong, the claim is accepted until future evidence repudiates the warrant.
- Rational thought, the evidence available, and the data collected (observed) shape knowledge at the time.
- Research is relevant and contextual. Post-positivist researchers seek statements that are relevant and true, given the context, to explain a causal relationship in the research.
- Post-positivists seek to be objective, basing research on the “best” evidence available at the time.

What one can note, under a post-positivist approach, is that the relative and contextual truth that is observed, given the nature of the research methods, may be the best knowledge at the moment. Phillips and Burbules (2000) stated, “Science does not attempt to describe the total reality (i.e., all the truths) about, say, a classroom; rather, it seeks to develop relevant true statements – ones that can serve to explain the situation that is of concern or that describe the causal relationships that are the focus of interest” (p. 38).

Often, the quantitative method uses strategies of inquiry based surveys or experiments. The instruments used in quantitative methods are designed to yield data for statistical analysis. Quantitative research also tends to be interested in whether and to what extent variance in one factor is related to variance in another factor (Creswell, 2003; Maxwell, 2005; Sansone, Morf, & Panter, 2003). Results from quantitative research can often be generalized, though this is not always the case.

Research Context

This study investigated the field of education technology leadership as it related to pre-kindergarten through twelfth grade (P-12) educators from an urban central Iowa school district. Eight Technology Leadership (TL) standards identified in the literature were used as the characteristics of education technology leadership of primary interest to the study. There was not a requirement in study for participants to have formal knowledge or training related to the TL Standards.

The primary research questions for this study were defined in Chapter 1 as:

- (1) What are the set of items aligned with each of eight education technology leadership standards that would be part of a reliable and valid survey instrument that would assess the education technology leadership skills of an individual who completes the survey?
- (2) What are the underlying structures of the data generated by these items?
- (3) What are the inter-relationships between the underlying data structures and the education technology leadership standards?

In response to the research question of this study, a quantitative instrument in the form of a survey, the Educators Technology Leadership Assessment (ETLA), was developed. The ETLA survey was designed to assess an individual's education technology leadership abilities. The goal in developing the ETLA was to produce a short, multiple-choice assessment designed to measure the school technology leadership of a respondent across the full range of the eight TL Standards.

This study was conducted in two main phases: (1) the ETLA Development Phase, and (2) the ETLA Field Test Phase. In the ETLA Development Phase, the research

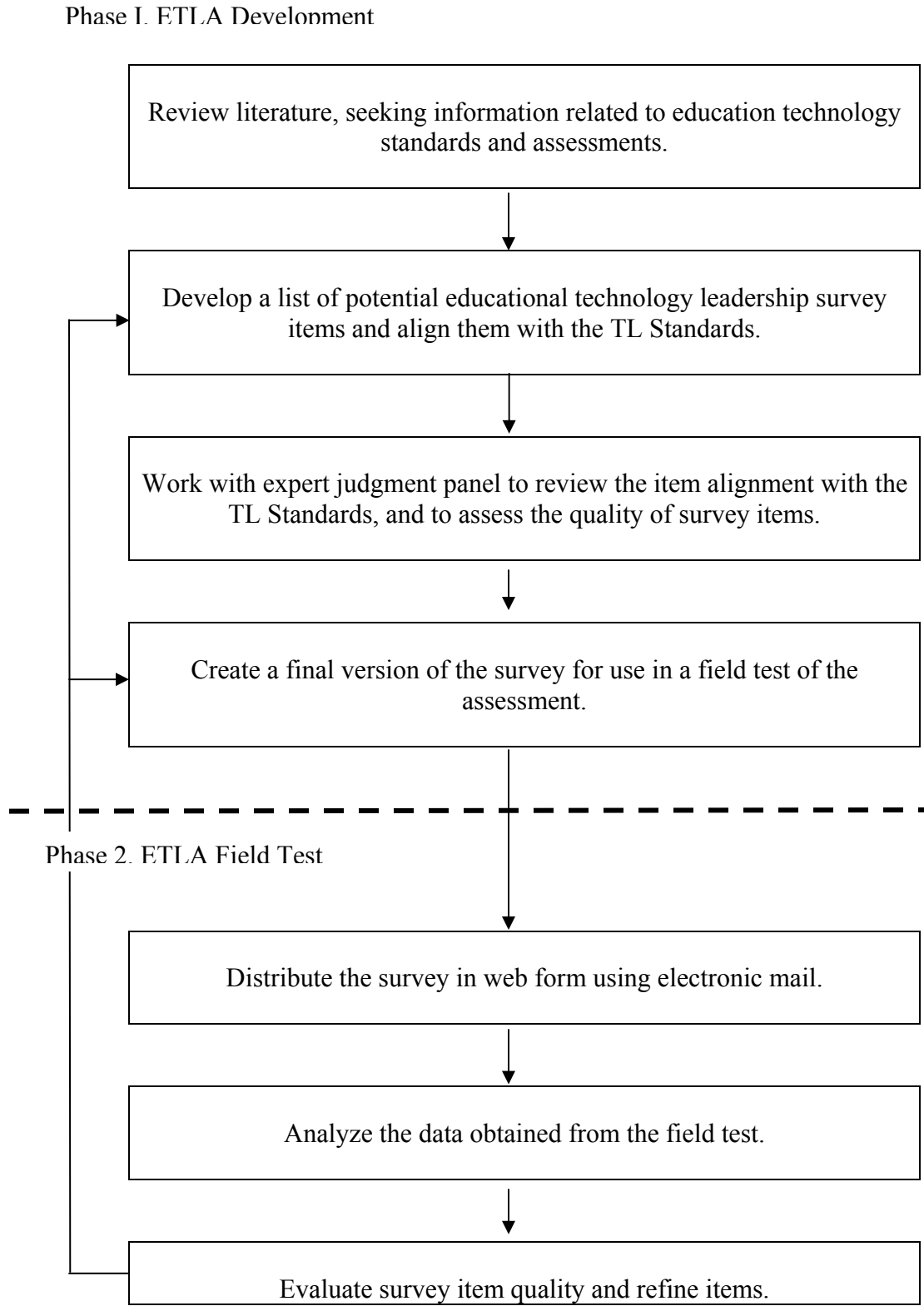
methods were primarily designed to support the creation of valid and reliable survey items. In the ETLA Field Test Phase, the research methods were primarily designed to test the reliability and validity of the survey items. Figure 2 shows the milestone events of the study across the two study phases, in chronological order.

ETLA Development Phase

Research Design Overview

The main task in the ETLA Development Phase was the construction of the initial draft version of the ETLA survey items. The draft ETLA survey items were constructed using a process similar to the method used by the American Institutes of Research (AIR) in developing the Principal Technology Leadership Assessment (PTLA) (Center for the Advanced Study of Leadership in Education, 2005). At the time of this study, the PTLA was generally available to P-12 school organizations and educational leadership preparation programs. The survey questions used in the PTLA were based on ISTE National Educational Technology Standards for Administrators (NETS-A) (ISTE, 2004). In comparison, for the ETLA, the ISTE publication *Teachers as Technology Leaders* (Twomey et al., 2006) was used as a guide by the researcher in the construction of ETLA survey questions. The ETLA survey items were designed to align with the existing Technology Leadership (TL) Standards that were documented in the ISTE guide (p. 152-170). Development of the instrument began with a review of the TL Standards by the researcher to identify specific behaviors, activities, and practices associated with each of the standards. The information gathered in the review was used to help the researcher draft forty-four ETLA survey questions.

Figure 2. ETLA survey development process.



A goal of the researcher in the development of the draft ETLA items was to optimize items so as to elicit reliable and valid responses. The researcher was able to accomplish some item optimization directly as draft ETLA items and survey instructions were created. For example, the researcher ensured that the draft ETLA items asked respondents about past behaviors (rather than intended behaviors). And ETLA survey instructions directed respondents to develop their responses based on a discrete period of time.

To improve the likelihood for ETLA items to be reliable and valid, it was important that the ETLA survey items contained a basic level of quality. Groves (2004) suggested three quality standards that survey questions should meet: (1) content standards (i.e., are the questions asking the right things?), (2) cognitive standards (i.e., do respondents understand the questions consistently?), and (3) usability standards (i.e., can respondents complete the questionnaire easily and as they were intended to?). The development of survey questions with good quality would contribute to the goal of obtaining reliable and valid questions.

In order to provide support to the researcher in the important process of quality survey item development, the use of an expert judgment panel (Bass & Avolio, 1997) was incorporated into the study's methods. The expert judgment panel was used to help the researcher evaluate the quality of the draft ETLA survey questions. The responsibility of the expert judgment panel for this task was to review draft ETLA items and provide feedback related to the quality of each item. The data set that resulted from the collection of this expert feedback was expected to provide data useful for assessing ETLA item quality related to Groves's content, cognitive and usability standards.

In addition to the development of items with good quality, a goal in the development of the ETLA was to create a tool that would assess education technology leadership in terms of the eight TL Standards. To accomplish this, it was necessary to align the ETLA items with specific TL Standards. Each of the forty-four draft ETLA questions were initially assigned to a TL Standard area by the researcher as they were created. The result of this first alignment was that each standard area had between five and seven questions associated with the standard area. The responsibility of the expert judgment panel for this task was to review draft ETLA items and provide feedback related to the alignment of ETLA items with the TL Standards. The information obtained from the expert judgment panel was reviewed by the researcher to determine if the ETLA items had been properly aligned with the TL Standards.

Research Participants

The main participants in the ETLA Development Phase were the members of the expert judgment panel. The expert judgment panel was convened as part of one regular meeting of twenty-eight technology coordinators from central Iowa (United States of America) school districts in April of 2007. Based on their direct job responsibility as district technology coordinators to provide technology leadership to their respective school districts, these individuals were considered to have expertise in the area of education technology leadership. The technology coordinators participated as experts in this study on a volunteer basis.

Phase Data Collection

Data associated with the draft ETLA survey items were collected through the work with the expert judgment panel. An important aspect of the data collection was the

use of data coding to assist in tracking data. It was anticipated by the researcher that ETLA survey items would be modified from their original draft form over the course of the study. In fact, over the course of the study, two main versions of the ETLA survey existed, the draft version and the final version. It was necessary to develop data coding methods that would assist the researcher in identifying the various versions of the ETLA items over the course of the study.

The draft ETLA version of the survey was used for work in the ETLA Development Phase of the study. In the draft ETLA item coding scheme, labels for items associated with the draft ETLA survey started with a D. The standard area that the item was aligned with was identified with a roman numeral (I-VIII). A number was used for the marker of the item within the standard area. For example, the item coded as D.V.3 identified question 3 in TL Standard area V (Productivity and Professional Practice) from the draft version of the ETLA survey.

The main data collection associated with the ETLA Development Phase was conducted with the expert judgment panel, as part of a meeting of technology coordinators. To accommodate meeting time constraints, the work of the expert judgment panel was designed to be accomplished in one two hour meeting of the panel. Given this time constraint, it was not possible for all items to be reviewed by all experts. Therefore, a process was designed to insure that each item would be reviewed by ten experts. The assignment of items to various experts for review had an element of randomness to it, in that no item was reviewed by the same ten experts.

At the meeting of the expert judgment panel, an activity designed to solicit feedback from the panel was completed. First, the researcher reviewed the TL Standards

with the entire panel. This allowed the experts to become more familiar with the TL Standard areas and to seek clarification about the TL Standards.

Next each expert was provided ETLA Item Expert Feedback forms. As previously indicated, the distribution of these forms was conducted so that each ETLA item would be reviewed by ten experts. A single ETLA Item Expert Feedback form contained one ETLA item, the TL Standard that it had been associated with by the researcher, and two general purpose requests for the expert to respond to, related to the item. First, the experts were asked to self-assess themselves by responding to the item using a six point scale. Second, the experts were asked to rank the item's alignment with the TL standard area the item was assigned to using a five point scale. An area was also provided on the form for the respondent to provide additional comments related to the item.

A copy of the expert feedback form for item D.I.1 is shown in Figure 3. Copies of the entire set of ETLA Item Expert Feedback forms are contained in Appendix B.

Two data sets were generated from the work with the expert judgment panel. One data set, labeled the Expert Ranking Data Set, contained data related to the ranking of the item's strength related to the item's TL standard area. The other data set, labeled the Expert Assessment Data Set, contained data related to the item's use as a self-assessment question. The data sets collected from the expert panel along with comments submitted on the feedback forms were used by the researcher to help analyze the quality of the ETLA survey questions.

Figure 3. Example of expert panel feedback form

Standard: Leadership and Vision.

Educational technology leaders will facilitate development of a shared vision for comprehensive integration of technology and foster an environment and culture conducive to the realization of the vision. Educational technology leaders:

- A. Identify and apply educational and technology-related research, the psychology of learning, and instructional design principles in guiding the use of computers and technology in education.
- B. Apply strategies for and knowledge of issues related to managing the change process in schools.
- C. Apply effective group process skills.
- D. Lead in the development and evaluation of district technology planning and implementation.
- E. Engage in supervised field-based experiences with accomplished technology facilitators and/or directors.

Question – D.I.1

To what extent did you participate in your district's or school's most recent technology planning process?

- Not applicable
- Not at all
- Minimally
- Somewhat
- Significantly
- Fully

How well do you believe this question measures this standard area?

- Not at all
- Weak
- Somewhat
- Strong
- Very Strong

Please provide any suggestions or comments regarding the question above in the space below. If you have suggestions for additional questions please write the questions in the space below.

Data Analysis

To aid the evaluation of the quality of the draft ETLA items, descriptive statistics were generated using the Expert Ranking Data Set and the Expert Assessment Data Set. Included in the descriptive statistics was the number of responses (N) obtained for each item along with statistics describing the distribution of responses for the items. Item mean scores were also generated for each item for each data set. It was expected that item mean scores for the expert panel would be high. Therefore, items with low mean scores were identified as items of potentially low quality. In this way, item mean scores were an important descriptive statistic used in the analysis of draft ETLA item quality.

In the case of the item mean scores based on the Expert Ranking Data Set, the ETLA item mean score analysis was used to identify the strength of alignment of each survey item with a TL Standard area. Items with item mean scores greater than or equal to 4 were considered to be adequately aligned with the standard area. Item mean scores less than 4 indicated the items that were considered by the experts to be poorly aligned with the standard area they were associated with. These items were flagged by the researcher for modification or removal from the survey.

Related to the item mean scores based on the Expert Assessment Data Set, the data analysis was used to identify the items that generated results tied to self-perceived education technology leadership strengths of the experts. Items with low mean scores based on the Expert Assessment Data Set indicated items that may not be generating accurate results related to expert education technology leadership skills. The items with low mean scores (< 4.0) were flagged by the researcher for modification or removal from the survey.

Summary

The ETLA Development Phase of the study was designed to solicit ETLA survey item feedback from an expert judgment panel. The expert panel feedback process was designed to provide evidence of the quality of all forty-four draft ETLA survey items. The expert review also helped to confirm that ETLA survey items were aligned with TL Standards. Using feedback from the expert review, the researcher constructed the study's final version of the ETLA survey for use in the Field Test Phase of the study. There were thirty-eight items in the final version of the ETLA.

ETLA Field Test Phase

Research Design Overview

Research methods designed to produce an ETLA survey of good quality were used by the researcher in the ETLA Development Phase of the study. However, it is often necessary to evaluate the survey's quality through the use of a field test (Peterson, 2000). Rea & Parker (2005, pp. 31-32) describe a field test as a small-scale implementation of the survey that addresses critical factors such as:

Survey clarity. Will respondents understand the question? Are the response choices sufficiently clear to elicit the desired information?

Survey comprehensiveness. Are the questions and response choices sufficiently comprehensive to cover a reasonable complete range of alternatives?

Survey acceptability. Is the survey of excessive survey length, or are survey questions perceived to invade the privacy of the respondents, or abridge ethical or moral standards?

Incorporating a field test into this study's research design provided an additional set of research methods for the researcher to use to further assess the quality of the ETLA survey. Field testing methods provided data that allowed for additional analysis of ETLA reliability and validity.

For this study, reliability was defined as the extent to which the items would yield the same results with repeated trials (Carmines & Zeller, 1979; Peterson, 2000). The reliability of an item is determined by its consistency or dependability in measuring whatever it is designed to measure. The greater the reliability of an item the smaller the likelihood of measurement errors with the item. There were statistical methods which could be applied to field test data to evaluate ETLA item reliability. The reliability statistical methods chosen for this study are described in more detail later in this chapter.

Consideration of ETLA item validity was less straightforward. There are many forms of validity defined in the literature related to survey item development, including criterion-related validity, predictive validity, content validity, convergent validity, divergent validity, and construct validity (Dillman, 2007; Fraenkel & Wallen, 1993; Groves et al., 2004; Newman & McNeil, 1998; Weiner, 2000). Across these multiple definitions, a common characteristic is that validity is concerned with the relationship between theoretical constructs and indicators. This concept of validity, which is associated with a type of validity known as construct validity, was useful to this study. Construct validity is concerned with the "extent to which a particular measure relates to other measures consistent with theoretically derived hypotheses concerning the concepts that are being measured" (Carmines & Zeller, 1979, p. 23).

The field test of the ETLA provided data useful in the analysis of the construct validity of ETLA items. The construct hypotheses and statistical methods chosen to evaluate ETLA validity are described in more detail in the field test data analysis section found later in this chapter.

Research Participants

The population for the ETLA Field Test Phase of the study started with approximately 3,000 P-12 educators (administrators and teachers) from a single Midwest school district. Given constraints in the study's time and resources, it was not possible to survey all of these educators. Rather than use pure randomized (probabilistic) selection in the creation of the study's sample population, a purposive non-probabilistic method was used that produced a convenience sample (Hinkle, Wiersma, & Jurs, 2003; Newman & McNeil, 1998; Peterson, 2000), in the hope of achieving a high participant response rate.

To generate the convenience sample, each of the sixty Principals in the school district was asked to submit a team of five representatives from their building to complete the ETLA survey. Each team's membership was required to include the Principal and the Building Technology Specialist. The Building Technology Specialist was a teacher in each building assigned the extra job duty of on-site technology support for the school. The Building Technology Specialist served as a liaison between the school and the district Technology Department. The team's membership was rounded out with three other teachers of the Principal's choosing. The Principal was told only that their five member team would be asked to complete an education technology related survey, and that it was important to select team members that could be counted on to complete the survey.

The study's timeline coincided with a technology mini-grant program underway in the school district. In order to receive a technology mini-grant in the amount of \$1,000, schools were required to complete several technology planning related tasks. The completion of the ETLA survey by the school's five member team was added to the list of tasks required in order for buildings to receive the technology funding. Incorporating the completion of the ETLA survey into the funding task list provided an incentive to each building's five member team to complete the ETLA survey exercise as directed.

The result of these convenience sample selection methods were that sixty teams, each with five members, were identified to participate in the field test. This generated a potential survey participant pool of 300 participants: 60 Principals, 60 building technology specialists, and 180 teachers. Several of the statistical analyses used in the study required 200 or more responses, so it was important to achieve a high response rate with the ETLA field test.

Data Collection

The field test of the ETLA was facilitated with the use of a web-based version of the ETLA survey. The survey was provided to participants during the 2006-2007 school year by sending survey instructions along with a link to the survey to participants via electronic mail. The survey window was open for a three week period, from May of 2007 into June of 2007. Reminder messages were sent to participants that had not responded after the first week, and again after the second week. While responses were tracked by the researcher for data grouping purposes, the researcher pledged anonymity to participants in that their individual responses would be kept confidential.

Two data sets useful for analysis were created in the field test of the ETLA. One data set, called the Field Test Data Set, was generated from the use of the ETLA as a self-assessment by the sample population. The other data set, called the Matched Data Set, was generated from the use of the ETLA as an evaluation tool by a subset of Principals that participated in the study.

In order to generate the two data sets, two sets of instructions were developed. One set was used for the general participants (N=275) who were asked to complete the survey as a self-assessment of their own education technology leadership skills. These instructions can be found in Appendix C. The other set of instructions were developed for a sub-group of Principals (N=25) who were selected to use the survey to evaluate the education technology leadership of a specific team member from their building. These instructions can be found in Appendix D.

With both data sets, basic ETLA scoring results for each question existed across of range of 0 to 5. This scale was not precise but did provide a measure for each ETLA item. For each ETLA item, a score closer to 5 represented an area of perceived strength for the respondent; a score closer to 1 represented an area of perceived need for the respondent. A score of zero in a respondent's data record indicated that the question was considered *Not Applicable* by that respondent.

The option of Not Applicable was included as a response choice for each ETLA survey item. The decision to provide a Not Applicable response choice was made by the researcher in order to help capture information related to participants' understanding (or lack of understanding) of an item (Dillman, 2007). The Not Applicable response provided participants the ability to "opt-out" of answering a question; i.e. if they didn't feel it was

related to their job or if they didn't understand the question (Groves et al., 2004).

Responses of Not Applicable were coded in the data set as a zero and were treated as missing data in the statistical analyses used in this study.

There were limits associated with this item scale. For example, lower scores may have represented respondent(s) personal knowledge, skill, and/or level of interest (Rea & Parker, 2005). Lower scores may also have reflected a lack of opportunity for involvement by respondents in the area of education technology leadership. A high number of Not Applicable responses may have reflected a lack of understanding by respondents of the standard area as it related to the use of education technology.

The field test data collection methods were designed to allow the data to be examined on both an individual and group basis. The data generated by the field test provided the ability to evaluate item responses, item scale responses, and the inter-relationships between items and items scales. For example, the item alignment with TL Standard areas formed item scales for the TL Standard areas. These scales were available for statistical analysis and interpretation; e.g., respondents' scores for the scale of items F.I.1, F.I.2, F.I.3, F.I.4, and F.I.5 were averaged to create a mean score for the respondents for TL standard area I (Leadership and Vision). Mean scores close to 5 represented overall strength for that standard area for the field test respondent(s). A score close to 1 represented overall an area of need for the respondent(s) for that standard.

The item mean scores also provided a basis useful in comparing different subgroups within the field test sample population. For example, the item mean scores for Principals, Building Technology Specialists, and Teachers could be averaged to form a group item mean score, and then compared. The group item mean scores close to 5

represented overall strength for that item for the group. A score close to 1 represented overall an area of need for the group for that item.

Data Analysis

This section will discuss the statistical analysis methods chosen to analyze the field test data. In addition to continuing the evaluation of ETLA survey quality that was started in the ETLA Development Phase of the study, the researcher selected methods expected to provide evidence of ETLA item reliability and validity, as defined previously in this chapter. The researcher also sought evidence of underlying data structures within the Field Test Data Set, using exploratory factor analysis research techniques.

ETLA Reliability Analysis

The Field Test Data Set was used in the evaluation of ETLA survey reliability. Evidence of survey reliability can be demonstrated using measures of internal consistency (Hinkle, Wiersma, & Jurs, 2003; Schmitt, 1996; Welkowitz et al., 2006). Cronbach's Alpha (Alpha) was used in this study as a common estimate of internal consistency of ETLA items. Alpha measured the extent to which ETLA item responses correlated highly with each other. The widely accepted social science cut-off is that Alpha should be .70 or higher for a set of items to be considered a reliable scale (Schmitt, 1996). A collection of items with a high Alpha score does not necessarily imply that the items are measuring only one construct. A set of items can have a high Alpha and still be multidimensional. This characteristic of Alpha was helpful in the analysis of ETLA field test data, as the ETLA was intended by design to have clusters of items aligned to the eight technology leadership standards.

Items with good internal consistency can generate a low Alpha score if there was high random error in the item responses (Hinkle et al., 2003). The study's reliability analysis using Alpha was limited by this risk. Also, the risk of random error in this study was increased due to the availability of only one field test data set for use with Alpha analysis. Also, it is worth noting the statistical formula for Alpha took into account the number of survey items. Typically, as the number of items increase for a scale, the higher the Alpha score. It would be expected in this study's results that the Alpha for the ETLA survey overall (38 items) would be higher than the Alpha for each TL Standard aligned subgroup of items (4-5 items).

ETLA Validity Analysis

For analysis of ETLA indicators of validity, both the Field Test Data Set and the Matched Data Set were used by the researcher. The researcher was interested in finding evidence of ETLA construct validity. As mentioned previously, construct validity is concerned with the extent to which a particular measure relates to other measures consistent with theoretically derived hypotheses concerning the concepts that are being measured. Construct validity seeks agreement between a theoretical concept and a specific measuring device or procedure. To understand whether a piece of research has construct validity, three steps should be followed (Carmines & Zeller, 1979, p. 39). First, the theoretical relationships must be specified. Second, the empirical relationships between the measures of the concepts must be examined. Third, the empirical evidence must be interpreted in terms of how it clarifies the construct validity of the particular measure being tested.

In order to examine data collected in the field test for evidence of ETLA construct validity, it was necessary to conduct experiments designed to assess construct validity. Three such assessments were developed: (1) the Job Group Assessment, (2) the Matched Data Assessment, and (3) the Not Applicable Assessment.

The Job Group Assessment was conducted by examining job groups. For the job group construct validity assessment, the hypothesized theoretical construct was: Valid ETLA items would differentiate the leadership across these job groups, and so the education technology leadership of each job group would vary compared to the others.

For this assessment, it was necessary to group the field test responses by job category. The field test sample population contained respondents from one of three job groups: (1) Principal, (2) Building Technology Specialist, and (3) Teacher. A claim for ETLA item construct validity would be supported if item mean scores for each job group varied.

The Matched Data Assessment was the second examination of ETLA construct validity. In this test, the Matched Data Set was generated by matching responses from selected principals with responses from selected teachers. To create the Matched Data Set, 25 of the 60 Principals in the field test population were selected at random to use the ETLA survey to evaluate the technology leadership of a randomly selected member of their building's team. Instead of using the ETLA survey as a self-assessment, these 25 Principals were asked to evaluate the technology leadership of a specific team member. In turn, the team member used the survey as a self-assessment of their education technology leadership skills. The 25 responses from Principals were compared with the matching (paired) 25 responses from team members.

For the matched data construct validity assessment, the hypothesized theoretical construct was: The scores of the principal would relate to the scores of the teacher. For this assessment, the indicators of primary interest to the researcher were (1) the ETLA item mean scores of principals compared to teachers, and (2) the correlation of the matched responses between principals and teachers. The a priori expectation was that a claim for ETLA item construct validity would be supported if item mean scores for Principals and Teachers were similar and if item scores were highly correlated.

The Not Applicable Assessment was performed by reviewing *Not Applicable* responses in the Field Test Data Set. All ETLA items offered study participants the opportunity to respond that the item was Not Applicable for them in their job. Because ETLA items had been aligned, with the aid of an expert judgment panel, to education technology leadership standards considered in fact to be applicable to educators, ETLA items with many Not Applicable responses indicated items with potentially low content validity. For the not applicable construct validity assessment, the hypothesized theoretical construct was: Valid ETLA items would be considered applicable by respondents. For the purpose of this study, ETLA items with a Not Applicable response rate greater than 5% of the total item responses were identified as items with potentially low construct validity.

ETLA Exploratory Factor Analysis

Factor analysis is a generic name given to a class of multivariate statistical methods whose primary purpose is data reduction and summarization (Hair, 1979). Factor analysis can be useful in supporting research methods designed to examine the relationship between variables within a data set. Tabachnick and Fidell (2001) stated:

Factor analysis is a statistical technique applied to a single set of variables when the researcher is interested in discovering which variables in the set form coherent subsets that are relatively independent of one another. Variables that correlated with one another but largely independent of other subsets of variables are combined into factors. Factors are thought to represent underlying processes that have created correlations among variables. (pp. 582-583)

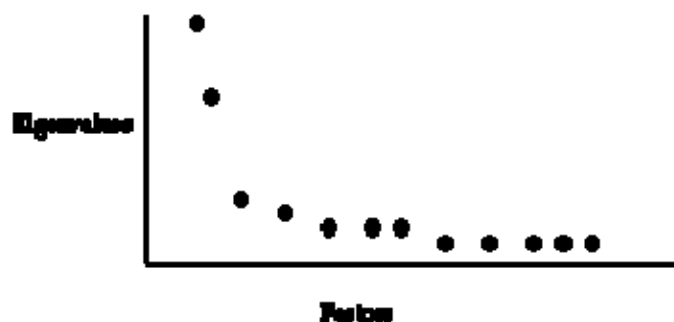
Exploratory factor analysis (EFA) is a form of factor analysis methodology that is generally used to discover the factor structure of a measure. EFA is used “as a means of exploring the underlying factor structure without prior specification of number of factors...” (Kim & Mueller, 1978b, p. 77). EFA is an “expedient way of asserting the minimum number of factors that can account for the observed covariation and as a means of exploring data for possible data reduction” (p. 9).

The first goal in EFA is to determine the number of factors. A common technique in EFA useful for determining the number of factors is to generate a Scree plot. The Scree plot is a two dimensional graph with factors on the x-axis and eigenvalues on the y-axis. Eigenvalues are produced by a process called principal components analysis (PCA) and represent the variance accounted for by each underlying factor. They are not represented by percentages but instead by scores that total to the number of items. For example, a 38-item scale, such as the ETLA, will theoretically have 38 possible underlying factors; each factor will have an eigenvalue that indicates the amount of variation in the items accounted for by the factor. If the first factor has an eigenvalue of 3.0, it accounts for 8% of the variance ($3 \div 38 = .08$). The total of all the eigenvalues will

be 38 if there are 38 items, so some factors will have smaller eigenvalues. They are typically arranged in a Scree plot. Figure 4 shows an example of a Scree plot.

As can be seen in the Scree plot example in Figure 4, the first two factors account for most of the variance and the remaining factors all have small eigenvalues. A researcher might examine the Scree plot and decide there are two underlying factors. This

Figure 4. Example of a Scree plot



approach to selecting the number of factors involves a certain amount of subjective judgment.

Another approach to deciding the number of factors is called the Kaiser-Guttman rule (Kim & Mueller, 1978a). This method simply states that the number of factors are equal to the number of factors with eigenvalues greater than 1.0. For this study, examination of a Scree plot along with the Kaiser-Guttman rule will be used to determine the number of factors found through EFA.

Once the number of factors is determined, the next goal in EFA is to describe the factors. A factor is often described in terms of the items that load high on the factor (Kim

& Mueller, 1978b). The factor loadings generated by the EFA are reported in a table called a factor matrix. The matrix can be examined to identify items that have high factor loadings on a specific factor. The grouping of items that load high on a specific factor is called the factor scale. For example, consider the sample factor matrix in Table 5. This matrix would indicate a factor structure of three factors (columns) and five items (rows). In this example, the items identified with high loadings are highlighted in bold type. Item 1 was found to load high on Factor 1 and Factor 3. Item 1 would be considered to be *multidimensional*. Item 3 loaded high on Factor 2, Item 4 loaded high on Factor 1, and Item 5 loaded high on Factor 2. These items would be considered unidimensional. Item 2 did not load high on any factor. In this example, the matrix would provide evidence indicating that Item 1 represents more than one factor, and that Item 2 should be evaluated to determine if it is contributing to the researcher's overall goals.

There are a wide number of options and settings available to researchers conducting factor analysis (Byrne, 1989; Carmines & Zeller, 1979; Kim & Mueller, 1978a, 1978b; Kline, 1994). The SPSS computer program (SPSS Version 11 for Windows) was used in this study to generate the statistics of the EFA. The SPSS software application was programmed by the researcher to use the Principal Axis Factoring option (Kim & Mueller, 1978b). The Principal Axis Factoring setting was appropriate for the factor analysis methods of this study, because this setting caused the SPSS program to perform multiple iterations using communalities estimates, until the changes in communalities from one iteration to the next satisfied the program's convergence criterion for extraction. In other words, the factors generated using Principal Axis Factoring would be expected to account for much of the variance in the data.

Table 5. Sample factor matrix.

Variable	Factor 1	Factor 2	Factor 3
Item 1	0.723	0.098	-0.905
Item 2	-0.250	-0.294	-0.314
Item 3	-0.028	-0.879	0.012
Item 4	-0.952	-0.380	-0.578
Item 5	0.297	0.925	-0.204

Other settings used by the researcher in this study's SPSS based factor analysis included:

- Kaiser-Meyer-Olkin measure – The Kaiser-Meyer-Olkin (KMO) setting tests whether the partial correlations among items are small. High values (close to 1.0) generally indicate that a factor analysis may be useful for the data set. If the value is less than 0.50, the results of the factor analysis would be less useful.
- Bartlett's test of sphericity – The Bartlett's Test of Sphericity tested the hypothesis that the correlation matrix is an identity matrix, which would indicate that the items are unrelated and therefore unsuitable for structure detection. Small significance level values (less than 0.05) would indicate that a factor analysis may be useful with the data set.
- Anti-image – The anti-image setting generated the anti-image correlation matrix and the anti-image covariance matrix. Small off-diagonal elements indicate a good factor model. Sampling adequacy for an item is displayed on the diagonal of the anti-image correlation matrix.

- Varimax method – This setting caused SPSS to generate a rotation component matrix using an orthogonal rotation method that minimizes the number of variables that have high loadings on each factor.

The results of the EFA provided information useful for the researcher's examination of the interrelationships in the field test data. Results of the factor analysis included a Scree plot and a rotated structure matrix which were used by the researcher to examine the relationships between ETLA items and the factors generated by the factor analysis. The Scree plot and the associated eigenvalues were useful in helping the researcher to decide on the number of factors to consider. The rotated structure matrix was used to describe the factors by identifying ETLA items that loaded high on a factor, the ETLA items that did not load high on a factor, and the ETLA items that loaded high on multiple factors. Of greatest interest to the researcher for this study were the ETLA items that did not load high on any of the factors. As these items did not contribute in a significant way to the EFA, they were flagged for future study to determine if they were contributing to the purpose of the ETLA.

In summary, exploratory factor analysis (EFA) was used in this study to explore the interrelationships of the ETLA items. By examining the factor loadings generated from the SPSS based factor analysis the researcher was able to identify underlying data structures inherent in the Field Test Data Set. It was also possible to identify ETLA items with low factor loadings. These items would then become subject to future review to determine if they were contributing in a meaningful way to the ETLA results.

Summary

The field test was conducted with a sample population of educators and was enabled with the use of a web-based survey. Two data sets were designed to be created as part of the field test, the Field Test Data Set and the Matched Data Set. The data sets were statistically analyzed by the researcher in order to produce evidence of ETLA reliability, validity, and data inter-relationships. Statistical measures used included ETLA item mean scores, ETLA item scale mean scores, Cronbach's alpha, and exploratory factor analysis (EFA).

Design Issues

Internal Validity

Measures of internal validity are often used to help a researcher ensure that the relationship between variables is unambiguous and defines the researcher's intended relationships (Fraenkel & Wallen, 1993). When a study is said to have adequate internal validity, it means that any relationship observed between two or more variables should be meaningful in its own right, rather than being due to other factors. There are a number of factors that can threaten the internal validity of a research study, and the research design should attempt to control for as many of these threats as possible.

Threats to the internal validity of this study's research design included participant characteristics, loss of participants, history, testing, and implementation. This section discusses these threats to the internal validity of this study.

Participant characteristics. Selection bias of participants is one of the most common threats to the internal validity of a study. The researcher must be aware of and control for differences in the characteristics of participants that might interfere with the

variables being studied. There was selection bias associated with both of the participant groups engaged by this study. For example, in the case of the expert judgment panel the expertise of the panel member was assumed, as all participants in the expert judgment panel had professional job roles associated with education technology leadership. In the case of the field test sample population, the sample was not a true random sample but a convenience sample. All participants were employees of the same school district.

Loss of participants. One of the most difficult internal validity risks to control for is the loss of participants (Fraenkel & Wallen, 1993). This study's interaction with participants was limited to interactions at specific points in time for each group of participants, so the issue of controlling for participants over an extended period of time did not constitute a risk. What did cause a threat, in this study, was the potential failure of participants to complete all or part of the survey. Some of this risk was mitigated by using statistical methods to replace missing data. Also, an incentive in the form of the release of a technology mini-grant tied to completion of the ETLA survey task was used in the field test to help ensure a high survey return rate. The use of an online, web base survey in the field test added to the convenience for participants to complete the ETLA survey task.

History. Historical events that occur during a research project, such as an event that occurs in the middle of a data collection process, may have an impact on the participants' responses. In this study one control that was helpful in countering a potential threat in this area was that the field test data collection occurred over a short period of time; specifically, from May 2007 to June 2007. During this time period, there were no "significant" historical events identified by the researcher that may have impacted the

responses given by participants who completed the survey later in the data collection window compared to participants who completed the survey early in the data collection window.

Testing. In some instances, participants may have an opportunity to practice taking a test, survey, questionnaire, or scale prior to the research or as part of the research project (e.g., pre-test, post-test design). This can result in what is known as a “practice effect.” Participants who have seen or participated in preliminary or pilot tests may affect the way the participant responds in the future. For this study, participants were exposed to the associated ETLA tasks only once, which served to minimize the risk of a practice effect.

Implementation. When one group of participants receives special treatment that is not part of the research study an implementation threat has occurred. In this study, the implementation was standardized for all participants so that no one received any preferential treatment.

External Validity

The concept of external validity is synonymous with generalizing (Fraenkel & Wallen, 1993)). The external validity of a research study depends on the extent to which the results can be generalized. Creswell (2003) noted, “A threat to external validity arises when experimenters draw incorrect inferences from the sample data to other persons, other settings, and past or future situations” (p. 171). The external validity of this study was limited in that the use of only one expert judgment panel and the use of only one sample population limited the ability to generalize the results of this study.

Other Issues

Small Item Scales. Analysis of the reliability of the ETLA item scales would be bolstered by additional items aligned with each TL Standard area. Analyses and conclusions based on a scale of 4 to 6 items must be approached with caution (Avolio & Bass, 2002). Generally, a sub-scale consisting of 20 or more items is necessary to draw conclusions about a dimension's reliability (Rea & Parker, 2005). Additional trials of the ETLA would be helpful to the evaluation of ETLA reliability.

Limited number of data sets and sample populations. The use of factor analysis in this study was limited by the availability of only one data set. Kline (1994) cautions that only replicated factors should be interpreted. He also advises that the interpretation of factors from item content is not evidence of validity beyond simple face validity; i.e. claiming a test measures something from its appearance.

Small response set in Matched Data Set. The statistical tests using data matched with principals and teachers were based on a maximum sample population of N=25 matched cases. This sample size was small and findings would require verification through additional use of the ETLA. Also, the construct hypothesis used in the Matched Data Set analysis was not intended to infer that the Principal score was the "true" score; i.e., the methods were not intended to establish criterion-related validity. The intention was to evaluate ETLA item construct validity using the hypothesis that the matched scores would be similar and correlated.

Use of an incentive in the field test. In the study's field test, the completion of the ETLA survey by building teams was one of several tasks required for the building to receive a technology mini-grant. This served as an incentive for participants to complete

the survey. But it also invited the potential of participant bias in the survey. For example, the Principal for each building may have selected teachers most likely to complete the survey, creating a team not necessarily representative of the building at large. There was also the potential that participants would answer in a way they thought the Technology Department would want them to.

The field test participants knew they were part of a study tied to funding for their building, which introduced a potential for a *Hawthorne effect* to the study's validity. The Hawthorne effect is a term used to describe the changes in a subject's behavior when he or she is aware that he or she is being observed. In essence, the Hawthorne effect tells us that people do not behave in the same manner when they know they are being watched as compared to when they do not know they are being watched.

Human Subjects Research

Appropriate and timely Human Subjects Research forms (Office of Research Assurances, 2006) were filed with the Iowa State University Committee on the use of Human Subjects in Research. A copy of the human subjects' approval letter can be found in Appendix E. Permission to conduct research in the Des Moines Public Schools was also obtained. A copy of the letter of approval from the district is found in Appendix F. Participation in the survey was considered implied informed consent.

Summary

The availability of a measurement tool to assess education technology leadership would be useful to both individuals seeking to become better education technology leaders, and also to individuals seeking to study and develop education technology leadership. The methods of this study were designed to develop an education technology

leadership assessment (ETLA) with the aid of an expert judgment panel, and then to field test the assessment. Data generated from the field test was used to examine the ETLA for indications of reliability and validity. Interrelationships within the field test data were also examined through the use of exploratory factor analysis methods.

The methods of the study were limited in several ways, which posed risks to the study's internal and external validity. Because the study engaged human participants, Human Subjects Research guidelines were incorporated into the study's design, and the guidelines were followed by the researcher.