Utilizing the Rasch Model to Develop and Validate Items for the Tacit Knowledge Inventory for Superintendents (TKIS)

Christian E. Mueller¹
Kelly D. Bradley²

¹ Assistant Professor of Educational Psychology, The University of Memphis. Current research interests include giftedness and talent development (over the lifespan), and measurement issues in psychology and education. Current research projects focus on exploring resiliency in gifted students, and developing assessments of complex cognitive abilities utilizing the Rasch measurement model. Correspondence should be addressed to Christian E. Mueller, The University of Memphis, Counseling, Educational Psychology, and Research, 100 Ball Hall, Memphis, Tennessee, 38152.

² Assistant Professor of Educational Policy Studies and Evaluation, The University of Kentucky. Current research interests are anchored in quantitative evaluation and measurement, with specific application to the math and science teaching profession, specifically supply and demand issues, including professional development and quality in the classroom. Correspondence should be addressed to Kelly D. Bradley, The University of Kentucky, 131 Taylor Education Building, Lexington, KY 40506-0001.
ABSTRACT

Tacit knowledge was originally introduced into the professional literature by Michael Polanyi (1946, 1966, 1976), and later made popular by Sternberg and his colleagues (e.g. Sternberg et al., 2000; Wagner & Sternberg, 1985). Tacit knowledge is best characterized as context-specific, procedural knowledge that is gained from personal experience, rather than through direct training or instruction. The purpose of the current study was to develop and validate items for use on an assessment tool designed to measure tacit knowledge in superintendents. Development of the Tacit Knowledge Inventory for Superintendents (TKIS) occurred in three phases, including two phases of piloting and subsequent Rasch analysis. Among the more interesting findings was empirical support for the presence of intrapersonal, interpersonal, and organizational tacit knowledge in superintendents (Nestor-Baker & Hoy, 2001). Limitations included lack of sufficient items measuring at all levels of ability.
Utilizing the Rasch Model to Develop and Validate Items for the Tacit Knowledge Inventory for Superintendents (TKIS)

Understanding leadership effectiveness has long been an interest for researchers in both psychology and leadership. Leithwood (1995) argues that it is necessary to understand the cognition of the leader to be able to measure effectiveness. In their conception of practical intelligence (Sternberg et al., 2000), Sternberg and his colleague have offered the concept of tacit knowledge as one such cognitive approach. While numerous tacit knowledge measures exist, including the Practical Knowledge Inventory for Nurses, (PKIN, Fox, 1994, 1999), the Tacit Knowledge Inventory for Managers (TKIM, Wagner & Sternberg, 1991), and the Tacit Knowledge Inventory for Military Leaders (TKML, Hedlund et al., 2003), to date, there has not been an inventory designed specifically for use with school superintendents (Nestor-Baker & Hoy, 2001). Because tacit knowledge is context specific, it requires the development of a new inventory each time it is explored in a new domain. The current study offers a first step in the development and validation of a tacit knowledge inventory for use with superintendents.

Theoretical Foundation

Tacit knowledge, by its very nature, is a difficult construct to define and measure (Ambrosini & Bowman, 2001). Even though first introduced into the professional literature by Polanyi (1966), it was not until the later work of Sternberg and his colleagues (Hedlund et al., 2003; Sternberg & Hedlund, 2002; Sternberg & Horvath, 1999; Sternberg, Wagner, Williams, & Horvath, 1995; Wagner, 1987; Wagner & Sternberg, 1985) that considerable effort was put forth to validate, define, and measure tacit knowledge using empirical methods. Currently, there are three recognized characteristics that make tacit knowledge unique; namely, that tacit knowledge is procedural in nature, that it is useful in the accomplishment of personally meaningful goals, and that it is gained through personal experience (Wagner & Sternberg, 1985). Even more, it is proposed that tacit knowledge manifests as the tacit knowledge of managing oneself, managing others, and managing tasks (Wagner; Sternberg & Wagner). In the context of the superintendency, Nestor-Baker and Hoy (2001) have classified tacit knowledge into three domains: intrapersonal, interpersonal and organizational.

Traditionally, three methods have been applied in the development of psychological, personality and educational inventories: psychometrics or classical test theory (CTT), item response theory (IRT), and generalizability theory (Fox, 1999). Wright (1997) and Devillis
(1991), among others, have suggested that educational and psychological researchers are most familiar with the principles of psychometrics or CTT, since the majority of research conducted in these fields is grounded in the early work of Karl Pearson and Sir Francis Galton. In CTT, a researcher typically develops a bank of test items based on the use of comparable measures and/or theory. A good example of this is in the development of new items for use on some standardized measure or assessment, such as the SAT or GRE. Whether using comparable measures or being guided by relevant theory, the goal in using both of these methods is the same: to develop items that are both reliable and valid. The main objective is to measure the underlying theoretical construct of interest (Bond & Fox, 2001). In this process, both construct validity of items and internal reliability is of utmost importance to the researcher. It is essential that all items measure the same latent construct (construct validity) and that these items all relate to one another, as well (internal validity) (Bond & Fox).

Both Bond and Fox (2001) and Fox (1999) note, however, that there are inherent flaws in using psychometrics or CTT to develop items for use on psychological, personality and educational measures. First, it is typically accepted that Likert-type scale items are at interval levels of measurement; thus, implying an equidistant measure. In theory, this view may indeed be true; however, in practice it is flawed as there is no systematic way of ensuring this. A second assumption of CTT is that each item is of equal importance to the overall measurement of the construct, thus implying no redundancy or overlap of items. Third, it is typically assumed that all respondents are able to interpret directions correctly, that all items are written clearly, and that the items connect to a single construct. And lastly, CTT requires complete records in order to make comparisons and to produce a single standard error of measurement. In actuality, all of these assumptions are unstable, and often problematic, in traditional survey research (Sampson & Bradley, 2003).

Item response theory (IRT) attempts to address these limitations by providing a mathematical model in which a probability can be calculated. McCamey (2002) noted, IRT promises to overcome circular dependency of CTT which is the situation, as described by Fan (1998), where the person statistic is item dependent and the item statistics examinee (person) dependent…the Rasch model allows tests to be constructed where the measure of a person’s ability is independent of the sample of items used and is independent of the norm group used to ‘calibrate’ the test (p. 4).
In the current study, a one parameter IRT model known as the Rasch model (Rasch, 1960) was utilized. The Rasch model addresses many of the inherent weakness of CTT because, as Bradley and Sampson (2005) explain, “Rasch analysis begins at the level of measurement, providing diagnostic information on the quality of the measurement tool, in addition to yielding a more comprehensive and informative picture of the construct under measurement as well as the respondents on that measure”.

The earliest traditional IRT models dealt only with dichotomously scored items but were later applied to polytomously scored items such as Likert-type scales (Thornton, 2002). IRT models can involve multiple parameter estimations, including level of ability (theta), difficulty of item (beta) and probability of guessing. The Rasch model estimates only a single parameter, a person’s ability, and how it interacts with an individual item’s difficulty or endorsability. The Rasch model mathematically models the relationship between the latent ability of an individual and the item difficulty as a probability. Thus, as a person's respective ability increases, their likelihood or probability of answering correctly on any select item also increases. From this description, the utility of this method in the development of measures of latent abilities, like tacit knowledge, becomes obvious (Bond & Fox, 2001; Fox & Jones, 1998).

Objective

Nestor-Baker and Hoy (2001) wrote, “Perhaps the highest priority for practical application of superintendent tacit knowledge research is in the training and development of superintendents” (p. 125). However, before researchers and practitioners can begin to explore and understand how tacit knowledge guides effective superintendent decision-making, a new inventory must first be developed. The framework in the current study builds on the previous work of many researchers, mainly in the areas of cognition, leadership, and education. It combines elements of the research agenda of Sternberg and his colleagues used in studying tacit knowledge, aspects of the work of Nestor-Baker and her colleagues to push tacit knowledge research into the area of educational leadership, specifically the superintendency, and lastly the efforts of Fox and her colleagues (Fox 1994, 1999; Fox & Jones, 1998) in the use of the Rasch model.

The current research was conducted as part of a larger effort to develop a ready-to-use, pencil-and-paper assessment designed to measure tacit knowledge in school superintendents. When completed, the final version of the Tacit Knowledge Inventory for Superintendents (TKIS)
will be a scenario-based assessment that measures intrapersonal, interpersonal, and organizational tacit knowledge (Nestor-Baker & Hoy, 2001). Further, each scenario will be grounded in the eight professional standards of the superintendency developed by the American Association of School Administrators (AASA) (Hoyle, 1993). Hoyle, Bjork, Collier, & Glass (2005) have suggested that these standards offer the most comprehensive framework for evaluating effective superintendent performance.

Method

Development of the TKIS occurred in three phases: (a) generating the initial scenarios and response options; (b) initial piloting, Rasch analysis, and reworking of items; and (c) final piloting and Rasch analysis of reconstructed items. For illustrative purposes, only the results of the final phase of the larger study will be presented here. See Mueller (2006) for further details pertaining to the earlier phases of the larger study.

The final version of TKIS is designed to be a scenario-based assessment comprised of 40 tacit knowledge items across eight scenarios. Scenarios on the TKIS are designed to represent situations that are typical in a superintendent’s day-to-day professional role, and are assigned according to the eight AASA standards (described below). Each scenario presents an issue that has five corresponding response options, which reflect varying degrees of effectiveness in dealing with the issue. Respondents are asked to rate each of the response options on a 1- to 7-point scale where 1 = extremely ineffective to 7 = extremely effective. Scoring is done by matching a respondent’s choice to a scoring protocol, calibrated using expert responses. Credit is assigned dependent upon how well each of the respondent’s choices match the “correct” answer from 0 = no points to 3 = full points. Appendix A presents the initial form of the TKIS and contains eight scenarios and 33 tacit knowledge items. All items and scenarios were analyzed and calibrated according to the expectations of the Rasch model (discussed below).

Sampling Frame

The sampling frame consisted of 176 practicing superintendents and 15 graduate students from two superintendent preparation programs located within the state of Kentucky. While the graduate students were permitted to participate in both phases of piloting, superintendents were not. If for example, a superintendent was involved in phase I, developing scenarios and responses, that precluded involvement at later stages of the study.

Sample
Here, the focus is on the final phase of the study where the sampling frame consisted of six graduate students and 29 practicing superintendents, resulting in 35 responses being included in the final Rasch analysis. During the final phase, there were 117 assessments sent out to both graduate students and practicing superintendents. Six of the eight assessments distributed to graduate students were completed and returned, yielding a return rate of 75%. These responses were included in the final Rasch analysis. For practicing superintendents, there were 109 assessments distributed (via mail); of which, a total of 19 were returned yielding a response rate of 17.4%. To improve the overall response rate for practicing superintendents, and to meet the minimum necessary requirements to produce stable results in Rasch analysis (Linacre, 1994), reminder cards (Dillman, 1978) were sent to superintendents who had not yet completed the survey by the requested deadline. This resulted in an additional 10 surveys being returned for a final response rate for the practicing superintendents of 26.6%. Overall response rate for the combined group was 32.1%.

Data Collection Procedures

The process of developing the items and scenarios for use on the TKIS was completed in three phases. Phase I consisted of development of scenarios by experienced superintendents. These superintendents were identified based upon their expertise in one of the eight American Association of School Administrators (AASA) standards. These superintendents were then asked to generate scenarios and response options according to the criteria outlined by the lead researcher. Scenarios were designed to fit into one of three categories of superintendent tacit knowledge: interpersonal, intrapersonal, or organizational (Nestor-Baker & Hoy, 2001). In addition, each scenario was organized according to the AASA professional standards for superintendents. The eight standards are: leadership and district culture, policy and governance, communications and community relations, organizational management, curriculum planning and development, instructional management, human resources management, and values and ethics of leadership (Hoyle, 1993). As discussed previously, these standards have been recognized by some (e.g. Hoyle et al., 2005) as providing the most comprehensive framework for evaluating effective superintendent performance. Figure 1 provides a graphic depiction of the tacit knowledge framework used to organize the scenarios.
For those readers not already familiar with Rasch analysis, it may be helpful to provide a more thorough discussion of Rasch analysis and how it was used to analyze the tacit knowledge items in the final phase of the study. What follows is a general overview of the Rasch model, and in particular the partial-credit Rasch model (Masters, 1982), followed by a description of the specific methods used to analyze and calibrate the tacit knowledge items in the current study.

Analysis

The Rasch model. The Rasch model (Rasch, 1960) is particularly suited for use in developing and validating measures of complex cognitive abilities, such as the case with tacit knowledge (Bond & Fox, 2001; Fox, 1999). Wright and Linacre (2001) define the Rasch model as “a family of mathematical formulae for the relationship between the probability of being observed in a particular category, and the difference between an individual's ability and an item's difficulty” (p. 824). According to the Rasch model, the relationship between the latent ability of an individual and how difficult an item is can be mathematically modeled as a probability. As an individual’s respective ability increases, the probability of answering correctly on any select item also increases. (Bond & Fox; Fox & Jones, 1998). This probability, referred to as the dichotomous Rasch model since responses are modeled as either “right” or “wrong” is expressed mathematically as $ln(P_{nij}/P_{nij0}) = B_n - D_i$, where $P_{nij}$ is the probability that person $n$ encountering item $i$ is observed in category $j$, $B_n$ is the “ability” measure of person $n$, $D_i$ is the “difficulty” measure of item $i$, the point where the highest and lowest categories of the item are equally
probable (Linacre, 2005).

The partial-credit Rasch model. The partial-credit Rasch model (Masters, 1982) is an extension of the dichotomous model and is used for polytomously scored items, such as the case with items on the TKIS. The partial-credit Rasch model is expressed as

$$\ln(P_{nij}/P_{n(j-1)}) = B_n - D_i - F_{ij} = B_n - D_{ij},$$

where $F_{ij}$ is the “calibration” measure of category $j$ relative to category $j-1$, the point where categories $j$ and $j-1$ are equally probable relative to the measure of the item (Linacre, 2005, p. 20). Fox (1999) suggested that the use of the partial-credit model is particularly useful in developing assessments of complex cognitive ability, where “graded judgments of performance can be scaled and subsequently assessed for quality of calibrations” (p. 342). Furthermore, she notes, “the partial-credit model can calibrate these ordered levels of performance that represent varying degrees of performance” (p. 342).

Winsteps software (Linacre, 2004 version 3.51) was used to analyze data from both graduate students and experienced superintendents. Mean responses from the expert prototype were used as the scoring protocol. Data were analyzed separately for intrapersonal, interpersonal, and organizational tacit knowledge. The decision to analyze the data separately for each type of tacit knowledge, rather than as a single construct was made for two reasons: theoretical rationale and the assumption of unidimensionality in using the Rasch model.

Wagner (1987) found some support for an underlying single construct of tacit knowledge, but he found stronger support for the notion that tacit knowledge actually has three separate dimensions: intrapersonal, interpersonal, and organizational. Wagner used both principal-components analysis and confirmatory factor analysis to provide support for this assumption. Most researchers who conduct work in the area of tacit knowledge follow this precedent and study tacit knowledge as having three separate subdimensions. This was certainly the case for Nestor-Baker and Hoy (2001) who demonstrated empirical support for this in their work with superintendents.

In addition to the theoretical and empirical support for tacit knowledge being operationalized in this way, there were also two additional reasons that the Rasch analyses were conducted separately for each type of tacit knowledge. First, the assumption of unidimensionality is an important aspect of using Rasch analysis because of the presumption that all items on an assessment are supposed to measure the same construct. If all items on the assessment were analyzed together, the Rasch model would not work and many items would appear to misfit
when that was not necessarily the case. Secondly, the goal of the study was to use Rasch analysis
to create a measure or ‘ruler’ for each of the three subdimensions of tacit knowledge already
described. Using these rationale, the decision was made to conduct the analyses as they appear in
the present study.

*Analysis Procedures*

*Diagnosing misfit.* In order to determine how well the data fit the expectations of the
Rasch model (in this case unidimensionality), Winsteps software produces both infit and outfit
how well each item fits with the construct it is intended to measure. Infit and outfit statistics are
reported in either mean-square (unstandardized) or standardized form, and both are influenced by
the pattern of responses on any particular item. Tennant, McKenna, and Hagell (2004) noted,
“When the observed response pattern coincides with or does not deviate greatly from the
expected response pattern, the items fit the measurement model and constitute a true Rasch
scale” (p. 23).

In Rasch measurement, each item and a person’s response to that item can be
mathematically modeled as a prediction or a probability based on the relationship between the
person’s ability and the item’s difficulty. According to the Rasch model, as a person’s ability
increases, their likelihood of scoring correctly on an item also should increase. That is, if the
expectations of the Rasch model are correct. As such, misfit statistics provide an indication of
how well both persons and items are fitting according to the expectations of the model or the
prediction. In developing items for use on the TKIS, infit and outfit statistics were used to
indicate how well the individual items were functioning (as measures) of each type of tacit
knowledge.

Infit statistics are inlier-sensitive or information weighted and are more sensitive to
unexpected patterns of responses located near a person’s or an item’s measure (i.e. where the
model would predict responses to be). Therefore, if “less able” persons are scoring correctly on
the more difficult items on an assessment, then this would not meet the expectations of the Rasch
model. Outfit statistics are outlier-sensitive, and are more sensitive to unexpected patterns of
responses far from a person or item’s measure. For example, an item would indicate a poor outfit
statistic if several persons with low ability answered correctly on a more difficult item. This
would violate the expectation of the Rasch model that persons with low ability have a lower
probability of answering correctly on a more difficult item. In this case, unacceptable outfit
statistics will typically indicate either guessing or carelessness on the part of respondents
(Linacre, 2002a).

Mean-square fit statistics indicate the amount of randomness in responses on any
particular item, and thus show the amount of distortion in measurement. The expected value for
the mean-square statistic (infit and outfit) is 1.0. Values that are less than 1.0 indicate that
observations or responses are too predictable, or indicate redundancy. In other words, the item
does not provide useful information (from a measurement perspective) above and beyond that of
the other items contained on the assessment. Tang et al. (2005) note that this typically happens
either when items are highly correlated or when item responses are dependent on one another.
Values that are greater than 1.0 typically indicate unpredictability in responses, such as with
items that do not seem to be measuring the same construct. This is usually the result of either
poorly worded or ambiguous items (Linacre, 2002).

Standardized fit statistics are simply t-tests that test the hypothesis of whether the data fit
the model (perfectly). Linacre (2003) notes “Since the Rasch model is a mathematical ideal, like
a Pythagorean triangle, we never expect to encounter empirical data that match it exactly [or
perfectly]” (p. 918). The standardized fit statistic is reported as a z-score and indicates
improbability of the data, or whether the data did fit the model perfectly. The expected value is
0.0. Values less than 0.0 indicate that the responses are too predictable, and values greater than
0.0 indicates lack of predictability. Standardized fit statistics are obtained by converting the
mean-square statistics to normally-distributed standardized ones by means of the Wilson-Hilferty
cube root transformation (Linacre, 2002a). Therefore, in the cases of both the mean-square and
the standardized fit statistic, the question being answered in the use of fit statistics is “How
useful are the data as measurement units for the construct in question, or, how well do the data fit
the expectations of the Rasch model?”

Typically, there are acceptable ranges that items can fall for both infit and outfit statistics
for the item to still be considered a useful measurement item. Wright and Linacre (1994) note
that because the Rasch model is a stochastic, or probabilistic model, the choice of which cutoff
scores to use cannot be done with absolute certainty. They note, “When is a mean-square too
large or too small? There are no hard-and-fast rules” (p. 370). Wright and Linacre also note that
the choice of cutoff values will depend upon the type of Rasch analysis conducted. But, they also
go onto suggest that “fit statistics are indicative, not absolute” (Linacre, 1990, p. 81). Linacre further points out that the use of critical values in statistics is very similar to the use of fit statistics in Rasch. In other words, the use of critical values to reject the null hypothesis in statistics is not an absolute either and that the use of \( p < .05 \) (or whatever the case may be) is simply the acceptable level of tolerance for making a Type I error. The difference in Rasch however, is that the use of fit statistics is only to identify problematic items rather than to make inferences based on the use of critical values as in the case of statistical analysis.

**Variable (person/item) maps.** Winsteps software also produces a variable map where individuals are mapped alongside items in order to provide a graphic depiction of how well that bank of items is targeting the ability-level of the sample. Figure 2 presents a sample variable map that was taken from the measure for organizational tacit knowledge.

![Variable (person/item) map for interpersonal tacit knowledge.](image)

*Figure 2. Example variable (person/item) map for interpersonal tacit knowledge.*
On the left side of the map, it can be seen that there are 39 participants (referred to by “ID#”) being measured on their level of organizational tacit knowledge by the items on the right side of the map. There are 13 items on this particular measure (Figure 2). By placing persons and items side-by-side in this manner, a determination can be made about how well the individual items on the measure are targeting the full range of ability for the participants being sampled.

Bradley and Sampson (2005) liken the variable map, also referred to as a person/item map, to a ruler,

In this case, the items (on the right) act as the measurement units to gauge the ability measures of the persons (on the left). In a well-targeted assessment, mean item and person measures should be approximately equivalent, the difficulty measures of the items should span at least the width of the ability measures of the persons taking the assessment and the items should be distributed such that they accurately measure all persons taking the test. When a group of persons fall in a space between item placements on the ruler, it is comparable to measuring the length of an object with a ruler where the units of measure have been rubbed out at the very length of the object; one could report that the length of the object is between certain units, but could not report the precise length. (p. 13)

From the Rasch measurement perspective, the variable map in Figure 2 has three pressing issues: a) There are not sufficient items near the top of the continuum to measure the highest levels of ability, b) there are gaps between items where certain individuals are not being adequately measured (such as with ID#8 & ID#11), and c) there are redundant items, or items measuring at the same level of ability (e.g. Items 2a and 7b), contained on the measure. In order to improve this measure as an assessment tool, more difficult items would need to be added that targeted the higher levels of ability, items that “filled in the gaps” would need to be added, and finally redundant items should be deleted in order to make the assessment more concise. This last point is important because one of the main benefits of using Rasch analysis is to create an assessment that contains the fewest number of items that still provides good measurement information.

Two final pieces of important information to understand about the variable map is that both person ability (left side of map) and item difficulty (right side of map) are listed on a continuum from least able to most able, and from least difficult to most difficult. Secondly, the map is labeled on both sides with the person and item means (M), and one and two standard deviations above and below the mean (labeled respectively with S and T). Ideally, the means should be roughly equivalent, but on the measure in Figure 2 it can be seen that the mean for
persons is almost one standard deviation above the mean for items, thus indicating that the sample in this case found this measure relatively easy.

To summarize, in examining the first round of pilot data for measures of intrapersonal, interpersonal, and organizational tacit knowledge, items were examined for misfit (infit and outfit), as well as determining how well each of the three measures targeted the ability level for each of the respondents. Those items that were deemed to be outside the acceptable ranges (i.e. .6 to 1.5 for mean-square) were highlighted for further review. Items that were outside of the acceptable ranges were either re-worded or deleted in order to improve the overall reliability of each measure. In addition, where there were significant gaps on the continuum of items, especially if there were gaps at either end of the continuum, items were re-worded to improve the distribution.

**Expert Profile**

In order to create the scoring protocol for the TKIS, an expert profile was developed. Fox (1994) created a similar expert profile in her study by taking the mean responses of items of her pre-selected expert panel. In that study, Fox was able to identify four individuals to serve as the expert panel. Fox also included three additional criteria in creating her scoring protocol: (a) three of the four experts had to be in agreement to include the item for piloting, (b) responses for the experts had to be on the same side of the 7-pt. continuum, and (c) responses had to be within two points of each other. For example, an item would be excluded from further analysis if only one or two of the experts were in agreement on that item, or if two experts rated the best response as a “3” (somewhat ineffective) and another rated the best response as “5” (somewhat effective). This would meet the condition of being within 2 points, but would not meet the criteria of being on the same side of the continuum, therefore this particular item would be eliminated from further analysis. This protocol of using mean responses to develop the expert profile follows previous suggestions by Sternberg et al. (2000).

In the present study, the expert profile was calculated by taking the mean responses of those superintendents who had initially generated the scenarios in a previous step. Of the eight superintendents who were re-contacted, two agreed to take the full version of the assessment (containing all 17 scenarios) in order to create the profile, therefore only two experts were available to provide the responses to be used as the scoring protocol. One of these individuals was named outstanding superintendent of the year in the state of Kentucky in 2003. Each of
these individuals also had at least 10 years of experience on the job. This last qualifier was important because of previous literature suggesting that at least 10 years of experience is needed to develop expertise in any field (e.g. Chi, Glaser, & Farr, 1988; Ericsson & Crutcher, 1990).

Results

In order to illustrate the utility of the Rasch model in developing items for the TKIS, only the results from the final phase will be presented and then discussed. The results presented in this phase reflect changes made to the items during the earlier phases of the study, applying the Rasch model, and then reworking items that were problematic.

Results of Final Phase

The goal of this phase was twofold: to gain feedback on items using an increased sample size (Linacre, 1994) and to develop a final version of the assessment. At the end of the final phase, there were a total of eight scenarios and 34 items included for piloting and analysis. As a result of the Rasch analysis conducted in phase III, one item (8d) was dropped due to continued poor fit statistics. The final version of the TKIS has eight scenarios and 33 items. The TKIS can be viewed in Appendix A.

Intrapersonal Tacit Knowledge

Item misfit. Four intrapersonal tacit knowledge items were re-analyzed in phase III of the study. Referring to Table 1, it is evident that all but one item fit the expectations of the Rasch model. Specifically, item 8d had a mean-square outfit statistic (.55), thus indicating continued possible redundancy with this particular item. Item 8d was not included on the final version of the assessment. Item 8b, which was added during phase II to replace a poorly fitting item on the intrapersonal tacit knowledge measure fit the expectations of the model. In this case, replacing this item from phase II, which read Decide to do nothing, was replaced by Turn the situation over to the Board of Education. The latter response functioned much better. The corresponding fit statistics for this item were all within acceptable ranges, thus indicating it fit the expectations of the Rasch model. The other two items also fit the expectations of the Rasch model from an empirical standpoint.
Table 1

*Fit Statistics for Intrapersonal Tacit Knowledge (Phase III)*

<table>
<thead>
<tr>
<th>Item</th>
<th>Infit MNSQ</th>
<th>Outfit MNSQ</th>
</tr>
</thead>
<tbody>
<tr>
<td>8b</td>
<td>1.29</td>
<td>1.34</td>
</tr>
<tr>
<td>8c</td>
<td>1.00</td>
<td>1.03</td>
</tr>
<tr>
<td>8a</td>
<td>.83</td>
<td>.87</td>
</tr>
<tr>
<td>8d</td>
<td>.76</td>
<td>.55</td>
</tr>
<tr>
<td>Mean</td>
<td>.97</td>
<td>.95</td>
</tr>
<tr>
<td>S. D.</td>
<td>.20</td>
<td>.30</td>
</tr>
</tbody>
</table>

*Variable map.* In examining the final variable map (see Figure 3), the distribution of items appeared to be good. The four items seemed to be measuring at differing levels on the intrapersonal tacit knowledge continuum. Referring to Figure 3, it is also apparent that this group of intrapersonal tacit knowledge items has sufficient variability, as well. There are gaps in the items, where there are not enough items included on the measure to target the entire range of the sample. There is also a lack of items at the top of the continuum. It can also be seen that all of the graduate students are highlighted in bold, ID#1 – ID#6 on the right side of the variable map in order to illustrate how these individuals did compared to the practicing superintendents. It is evident that some of the students did quite well on this measure and some of the practicing superintendents did not. It is reasonable to think that these findings could be unique to the individuals sampled, or it could reflect that some of the practicing superintendents have not yet acquired a high level of intrapersonal tacit knowledge.
Figure 3. Variable (person/item) map for intrapersonal tacit knowledge (phase III).

Reliability estimates (person and item). Reliability in Rasch measurement, as in traditional statistics, indicates “the degree to which test scores are free from measurement error” (Smith, 2004a, p. 94). Similar to their use in true score theory, reliability estimates range from zero to one and are reflective of the inter-item correlation among all of the responses for a particular item, and thus are used as a measure of internal consistency for scores on that item.

Winsteps reports reliability estimates for both persons and items. Linacre (2005) notes,
Person reliability is equivalent to the traditional “test” reliability. Low values indicate a narrow range of person measures or a small number of items. To increase person reliability, test persons with more extreme abilities (high and low), [or] lengthen the test. Item reliability has no traditional equivalent. Low values indicate a narrow range of item measures, or a small sample. To increase item reliability, test more people [of varying abilities].

Furthermore, Bradley and Sampson (2005) note

Rasch measurement models provide a direct estimate of the modeled error variance for each estimate of a person’s ability and an item’s difficulty, providing a quantification of the precision of every person measure and item difficulty which can ‘be used to describe the range within which each item’s ‘true’ difficulty or person’s ‘true’ ability falls (Smith, 2004a, p. 96)” (p. 24).

Person reliability is improved when there is a large enough sample size (i.e. greater than 30, Linacre, 1994) and there are sufficient numbers of items to target the full range of abilities being measured; whereas, item reliability estimates are improved when there are sufficient numbers of individuals with varying degrees of ability on the latent construct being measured, including having few or no gaps in items where persons are not being measured.

Reliability estimates for intrapersonal tacit knowledge reflected the small number of items included in the final analysis (i.e. only four items), as well as the limited range of ability that was measured. In this case, including more items would probably improve the overall person reliability (.22) on the measure as more levels of ability would be able to be tapped or measured. The overall item reliability (.91) was deemed acceptable for this measure.

**Interpersonal Tacit Knowledge**

*Item misfit.* Seventeen interpersonal tacit knowledge items were re-analyzed in phase III of the study using the Winsteps software. Referring to Table 2, it is evident that all items that were re-analyzed and were included on the interpersonal tacit knowledge measure fit the expectations of the Rasch model. Thus, the four items that were re-worded (from phase II) and included as part of the analysis in phase III now fit the expectations of the Rasch model. Specifically, items 2e, 7b, 7d, and 7e now had fit statistics within the acceptable limits. Item 5 was added to replace an item that had poor wording in phase II. Items 7b, 7d, and 7e were reworded to remove extraneous information, which seemed to improve the understandability of the items. For example, the wording of item 7b in phase III…
You decide leadership training should begin with the board of education since they are to be held accountable for attainment of district goals.

Had been reworded from its phase II form of…

The board of education is to be held accountable for all goals attainment in the district. Therefore, the primary focus for leadership training needs to begin with the board of education albeit they are removed from day to day classroom activities.

In phase II, there was a lack of expert agreement on the item; therefore, the wording was changed which not only improved agreement between the experts, but also produced acceptable fit statistics in the final Rasch analysis.

Table 2

Fit Statistics for Interpersonal Tacit Knowledge (Phase III)

<table>
<thead>
<tr>
<th>Item</th>
<th>Infit MNSQ</th>
<th>Outfit MNSQ</th>
</tr>
</thead>
<tbody>
<tr>
<td>2d</td>
<td>1.24</td>
<td>1.26</td>
</tr>
<tr>
<td>2b</td>
<td>1.09</td>
<td>1.22</td>
</tr>
<tr>
<td>6d</td>
<td>1.16</td>
<td>1.18</td>
</tr>
<tr>
<td>6b</td>
<td>1.12</td>
<td>1.09</td>
</tr>
<tr>
<td>3a</td>
<td>1.08</td>
<td>1.07</td>
</tr>
<tr>
<td>7e</td>
<td>1.07</td>
<td>1.07</td>
</tr>
<tr>
<td>2a</td>
<td>1.06</td>
<td>1.06</td>
</tr>
<tr>
<td>7d</td>
<td>1.01</td>
<td>1.03</td>
</tr>
<tr>
<td>3c</td>
<td>.96</td>
<td>1.01</td>
</tr>
<tr>
<td>7b</td>
<td>.93</td>
<td>.99</td>
</tr>
<tr>
<td>2c</td>
<td>.95</td>
<td>.98</td>
</tr>
<tr>
<td>6c</td>
<td>.97</td>
<td>.92</td>
</tr>
<tr>
<td>3e</td>
<td>.97</td>
<td>.96</td>
</tr>
<tr>
<td>2e</td>
<td>.93</td>
<td>.86</td>
</tr>
<tr>
<td>6e</td>
<td>.85</td>
<td>.79</td>
</tr>
<tr>
<td>7c</td>
<td>.83</td>
<td>.79</td>
</tr>
<tr>
<td>3b</td>
<td>.76</td>
<td>.72</td>
</tr>
</tbody>
</table>
Variable map. In examining the final variable map for interpersonal tacit knowledge (Figure 4), several things become evident. First, the distribution of items on the interpersonal tacit knowledge measure seems to have sufficient distribution. In other words, the items seem to be measuring at various levels of tacit knowledge. Additionally, there is sufficient variability in the items, as evidenced by a wide distribution over the continuum (i.e. spanning up to 2 standard deviations (S and T) above and below the mean (M)). There continues to be, however, a lack of sufficient items near the top of the continuum where the most able individuals are. The most able respondents on this measure are not being adequately measured by this bank of items.

Referring to Figure 4, the graduate students are again highlighted in bold (ID#1 – ID#6) on the right side of the variable map to illustrate how these individuals did compared to the practicing superintendents. Again, it is evident that some of the students did well on the measure for interpersonal tacit knowledge as compared to some of the practicing superintendents. This could again reflect the notion that some of the practicing superintendents have not yet acquired a high level of interpersonal tacit knowledge, or it could be unique to the individuals in this sample.
Similar to the measure for intrapersonal tacit knowledge, the measure for interpersonal tacit knowledge demonstrated low person (.31) and item (.70) reliability estimates. Referring again to Figure 4, it is apparent that not all levels of ability are being measured (i.e. near the top of the continuum), as well as they could be. This also impacts overall item reliability, as there is not a sufficient range of ability levels (i.e. as evidenced by most people being gathered around the mean). Linacre (1994) notes that measuring individuals who have a wide range of ability on the construct of interest will improve the item reliability estimate, because the items are tapping into a larger distribution.

**Organizational Tacit Knowledge**

**Item misfit.** Thirteen organizational tacit knowledge items were re-analyzed in phase III of the study using the Winsteps software. Referring to Table 3, it is evident that all items that were re-analyzed and were included on the organizational tacit knowledge measure fit the expectations of the Rasch model. Thus, the two items that were shown as having being misfitting items in phase II now fit the expectations of the Rasch model in phase III. Remember that Item 4e in phase II (identified as item 1c in phase III) had a mean-square outfit statistic (.55) that was just under the acceptable threshold of .6. After reviewing the item, it was decided not to alter the
wording on this item because the wording seemed clear and understandable. It appears that rewording this particular item improved it in terms of its measurement properties, as the item fit the expectations of the Rasch model in phase III. Item 11d in phase II however did not improve and in fact was eliminated prior to conducting the Rasch analysis (i.e. eliminated due to poor expert agreement). Remember that in phase II this item was highlighted for further review, because it had a mean-square outfit statistic (.54) and in reviewing the wording it was determined to be poor. It is clear however, in examining results from phase II (i.e. poor fit statistic) and phase III (i.e. lack of expert agreement) that this was a misfitting item, which for whatever reason could not be improved. In phase II, the item had been worded…

*You call the assistant superintendent and district professional development committee together to share the concerns from staff members who are upset about the training.*

In phase III, the item was reworded…

*You share the concerns of upset staff members with the assistant superintendent and district professional development committee.*

Thus, extraneous wording was removed, but this still did not improve the item.

Table 3

*Fit Statistics for Organizational Tacit Knowledge (Phase III)*

<table>
<thead>
<tr>
<th>Item</th>
<th>Infit MNSQ</th>
<th>Outfit MNSQ</th>
</tr>
</thead>
<tbody>
<tr>
<td>5a</td>
<td>1.41</td>
<td>1.41</td>
</tr>
<tr>
<td>1c</td>
<td>1.26</td>
<td>1.26</td>
</tr>
<tr>
<td>1b</td>
<td>1.16</td>
<td>1.16</td>
</tr>
<tr>
<td>5c</td>
<td>1.06</td>
<td>1.06</td>
</tr>
<tr>
<td>5b</td>
<td>1.04</td>
<td>1.04</td>
</tr>
<tr>
<td>4a</td>
<td>1.04</td>
<td>1.04</td>
</tr>
<tr>
<td>1e</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>1d</td>
<td>.99</td>
<td>1.00</td>
</tr>
<tr>
<td>5e</td>
<td>.91</td>
<td>.93</td>
</tr>
</tbody>
</table>
In examining the final variable map for organizational tacit knowledge (Figure 5), several things become evident. First, the items on the organizational tacit knowledge measure seem to have sufficient distribution. In other words, the items seem to be measuring at various levels of organizational tacit knowledge. It is clear however, that there do remain gaps in this distribution (indicated by ellipse in Figure 5). Additionally, there is sufficient variability in the items, as evidenced by a wide distribution over the continuum (i.e. spanning up to 2 standard deviations (S and T) above and below the mean (M)). Contrary to the findings from the other two measures of tacit knowledge, there do appear to be sufficient items targeting the highest levels of ability in this sample. Thus, the most able respondents on this measure are being adequately measured by this bank of items.

Lastly, in comparing the performance of the graduate students to that of the practicing superintendents, it is evident in Figure 9 that more of the practicing superintendents are doing better than the graduate students. Whether this is a reflection of the sample or the nature of tacit knowledge would require further empirical study to determine.
Reliability estimates (person and item). Compared to the other two measures, the measure for organizational tacit knowledge demonstrated a very low person (.00) ability estimate, but an acceptable item (.83) reliability estimate. Referring again to Figure 5, it seems that the range of ability level is relatively small. Perhaps, measuring a wider range of ability would vastly improve the person reliability index. Item reliability, in this case, is at an acceptable level. It is apparent in reviewing Figure 5 that there are enough items measuring various levels of ability to produce an acceptable index in this case.

Discussion

Tacit knowledge is a form of procedural knowledge that is gained primarily from personal experience, rather than through formal training or instruction. Although it has been linked to on-the-job experience, it is more than simply accumulated years on the job. What appears to be more important than simply acquiring tacit knowledge, is in utilizing it to accomplish personally meaningful goals (Wagner & Sternberg, 1985). In this sense, tacit knowledge acts as a form of “meta-procedural knowledge” in that it operates at the cognitive
level and helps individuals balance many competing demands in their pursuit of these goals. Thus, it is more than simply procedural knowledge, because often it is the balancing of many different procedures that may determine successful outcomes. Lastly, tacit knowledge has been explored in the context of many professions, but it is usually defined as intrapersonal (managing self), interpersonal (managing others), or organizational (managing tasks). Because tacit knowledge is thought to be context specific, it typically requires the development of a new measure before it can be explored in the context of a new profession.

Nestor-Baker and Hoy (2001) conducted some preliminary work in this area when they explored tacit knowledge in the context of the superintendency. Nestor-Baker and Hoy found evidence that supported the work of previous researchers who found that tacit knowledge does take the form of intrapersonal, interpersonal, and organizational. In addition, they further posited that tacit knowledge could be broken down even further, or organized according to different characteristics. Intrapersonal, interpersonal, and organizational tacit knowledge were further broken down into 21 smaller categories. For example, Building and Sustaining Board Relations was characterized as interpersonal tacit knowledge. Previous researchers have also found that different types of tacit knowledge are more important for success depending upon the context. For example, Nestor-Baker and Hoy found that with school superintendents intrapersonal tacit knowledge is not as important for superintendent success as interpersonal tacit knowledge. Therefore, the final version of the TKIS reflects this finding in that there are more scenarios contained on the assessment for interpersonal than intrapersonal tacit knowledge. Although Nestor-Baker and Hoy’s study identified the types of tacit knowledge present in the superintendency, an assessment was never developed (Nestor-Baker & Hoy, 2001).

Thus, the current study was a first attempt to develop a ready-to-use, paper-and-pencil assessment that measured tacit knowledge in school superintendents and was primarily exploratory in nature. The goal was to develop an assessment that was based on solid theoretical rationale, as well as being an effective measurement tool (according to the principles of the Rasch model). While exploratory in nature, the results of the current study do offer promising insight into several important questions related to the development of an inventory that specifically measures tacit knowledge in the context of the superintendency.

*How well does the TKIS Measure Tacit Knowledge in School Superintendents?*
In reviewing the Rasch output for each of the three measures of tacit knowledge, there are different conclusions that can be drawn about how well each measure actually taps into the construct of tacit knowledge. While individual aspects of each measure will be discussed in more detail below, there are some summary statements that can be made about all three measures. First, it appears in reviewing the fit statistics for each measure, that the changes made to items based on the results of the Rasch analysis conducted in phase II, had a positive impact in phase III. For example, in phase II there were a total of 11 tacit knowledge items that did not fit the expectations of the Rasch model. On the final version of the TKIS (Appendix A), there was only one item that fell outside of acceptable limits. Secondly, in reviewing the variable maps, the items on all three measures appear to be bunched together around the mean. This indicates that there is not much variability in item difficulties, or, there is not a very wide range of abilities being measured. There appears to be improvement however, in the matching of item difficulty to person ability between phases II and III of the study. It could be that by including all items on all of the assessments the chance of getting easier items was reduced. Lastly, in reviewing person and item reliability estimates for each measure, it appears that while item reliabilities are at least acceptable, person reliabilities are poor in all cases. Again, this was probably influenced by the limited range of ability measured on each aspect of tacit knowledge. Linacre (2005) suggests measuring a wider range of ability in order to improve the person reliability estimate (addressed in further detail in Limitations and Future Directions). Next, details about each of the three measures of tacit knowledge are explored in-depth.

Intrapersonal tacit knowledge

Several conclusions can be made about the four items contained on the measure of intrapersonal tacit knowledge. In this case, only one of the items (8d) continued to be outside of acceptable limits of the Rasch model. The outfit mean-square statistic for this item was (.55), thus placing it below the .60 limit and possibly indicating observations were too predictable (Linacre, 2005). Reasons for this are not immediately evident and would need further investigation before this item would be considered an effective measurement item. Interestingly, this item shows up as being the “easiest” item on the continuum, one in which every respondent but one was able to get it correct. Secondly, the fact that items had adequate variability was a positive finding in that these items were clearly measuring at different levels of ability for this sample. However, it is also clear that more items are needed to adequately measure a full range
of ability in this area. Although intrapersonal tacit knowledge is not as important to superintendent success (Nestor-Baker & Hoy, 2001), the TKIS would be improved as an assessment tool by adding at least one more scenario (with five response options) to better measure the full range of abilities. Additionally, the inclusion of more items may also improve the person reliability estimate (.22) as more items will provide more information about more of the respondents.

Interpersonal tacit knowledge

Reviewing the results of the final Rasch analysis for interpersonal tacit knowledge provides some interesting feedback for these 17 items. First, fit statistics for all of these items are within the acceptable limits, thus indicating they are unidimensional or are measuring the same construct. Secondly, while there does seem to be some variability in the items, they are all clustered within +1 to -1 logits of the mean. There is also redundancy is some of the items, in that there are several items that measure at the same level of ability. There is also a clear need to have more difficult items included on the interpersonal tacit knowledge measure in order to better gauge at higher levels of ability. Lastly, person reliability is rather low (.31), but this would probably be improved if a more diverse sample was measured in terms of ability level. Item reliability was also relatively low (.70) and this again may be improved by adding more difficult items to the assessment.

Organizational tacit knowledge

The measure for organizational tacit knowledge also had all of the final items fitting the expectations of the Rasch model. Again, the adjustments made to the items between phases II and III of the study seemed to improve the items on this measure. The final 13 items all seemed to be functioning as measures of a unidimensional construct. Secondly, although there was some redundancy in which several items were measuring at the same level of ability, the items on the measure for organizational tacit knowledge targeted at the full continuum of ability in that items covered the full range, even those individuals with the most ability. In addition, the reliability estimates are also quite different. Where the reliability estimate for items is acceptable (.83), the person reliability estimate (.00) is concerning. Essentially, this indicates that a very narrow range of ability is being measured; thus, in order to improve the person reliability estimate, additional participants at various levels of ability would need to be measured.
In conclusion, the answer to the question “How well does the TKIS measure tacit knowledge in school superintendents” is at least partially answered by the efforts of this study. As a whole, there is only one item (8d) that did not function according to the expectations of the Rasch model. Therefore, it can be implied that the remaining 33 items are measuring the theoretical constructs from which they were developed. This probably reflects a combination of the efforts that went into creating the scenarios and items in phase I with using the results of the Rasch analysis to edit or drop items from further consideration in phase II. Secondly, as far as sufficiently targeting at the entire continuum of ability on all three measures, there is room for improvement. In the case of intrapersonal tacit knowledge, the limited number of items leaves significant gaps where many participants are not being adequately measured. For the measures of interpersonal and organizational tacit knowledge, there are seemingly enough items included on the measures. There is however, still some redundancy in items on both measures, and in the case of the measure for interpersonal tacit knowledge there are not sufficiently difficult items included on the measure to tap the higher levels of ability. Lastly, there is clearly room for improvement in both the person and item reliability indexes for all three measures. In this case, the addition of participants with a wider range of ability levels would certainly allow the clearer distinction of how well the individual items were functioning as measurement items. Because of the limited range of ability, and the assumption that tacit knowledge develops with more experience, the sample used here does not provide an adequate test for distinguishing this fact.

Additionally, the item reliability estimates would certainly be improved with the addition of items where there are currently gaps and also the inclusion of more difficult items on the intrapersonal and interpersonal tacit knowledge measures to better measure these higher ability individuals. On the whole however, it appears that the efforts at each phase of the study drastically improved items included on the final measure as indicators of intrapersonal, interpersonal, and organizational tacit knowledge.

Limitations, Strengths, and Future Research

Limitations

Limitations of the current study are best characterized as methodological issues, and concern the sample used, the limited number of experts used to generate the scoring protocol, and the very nature of the scenarios and tacit knowledge items generated in phase I of the study. Limitations in sample include the limited sample size used during phase II of the study, the
limited number of graduate students (novices) available during both phases of piloting, and the relative homogenous nature of the superintendents used. Linacre (1999) notes that with polytomously scored items, that as few as 10 observations per response offers at least reasonably stable information. However, Linacre (1994) and Wright and Tennant (1996) suggest using at least 30 responses per response in order to gain more useful and stable estimates.

Secondly, in regards to the limited number of graduate students available at both phases of piloting, this would obviously impact attempts to determine how well the items differentiate between different levels of ability. As such, the relatively small distribution of person ability (on the variable maps) was certainly impacted by the lack of graduate students. In addition, the poor person reliability estimates were also impacted by this fact. In future research, this fact will need to be addressed in order to gather more useful data on the distribution of ability on all three types of tacit knowledge. Lastly, in regards to sample, is the homogenous nature of the sample used in the present study. Again, it is difficult to ascertain whether within the state of Kentucky there exists the type of superintendents available to measure for the highest levels of expertise. This is certainly not to say that they are not available or even present, but rather that in the present study not enough effort went into establishing the criteria necessary to verify such high levels of expertise. In essence, the criteria of at least 10 years on the job (Ericsson & Crutcher, 1990) and being reputationally successful are not adequate to define the highest levels of superintendent performance and thus the theoretically highest levels of developed tacit knowledge.

The second notable limitation in the current study is that of generating the scoring protocol used to assign point totals for each response. Fox (1994) used the responses from four experts to generate her scoring protocol, and then utilized agreement between at least three of the four as an additional type of quality control. While efforts were made to contact more than the two experts utilized here, the reality is that only two agreed to this aspect of the study. Clearly, future research in this area should address this limitation by ensuring that more responses are utilized to generate the scoring protocol. In spite of this fact, it should be pointed out that two experts agreeing on an item still provides some validity to the understanding of what is the “best” way to handle a situation.

The final limitation discussed in this section concerns the nature of how scenarios and tacit knowledge items were generated during the first phase of the study. While previous researchers have included some form of interview protocol with their development of a tacit
knowledge inventory (e.g. Hedlund et al., 2001; Wagner & Sternberg, 1985), this was not done in the current study. In order to maximize the likelihood of respondents agreeing to participate in this phase of the study (i.e. generating scenarios), the decision was made to include explicit instructions with the information packet sent to each person. In retrospect, it seems that this protocol with the addition of interviewing with a limited number of participants would have allowed the primary investigator to decipher ambiguous responses, etc. Future researchers in this area would likely improve their development of inventories of this type by including this important part of the process.

**Strengths**

This study has three notable strengths that are worth highlighting. First, it is the first attempt to develop a ready-to-use, paper-and-pencil tacit knowledge inventory for use exclusively with superintendents. Secondly, development of scenarios and items were based on the eight professional standards of the American Association of School Administrators, which represent acceptable standards of practice in this area. Lastly, through the use of the Rasch model, and with the inclusion of two rounds of data collection and subsequent analysis, the items that were included on the final version of the TKIS should be considered unidimensional items that represent intrapersonal, interpersonal, and organizational tacit knowledge (Smith, 2004b). Thus, while the efforts in the current study did not produce a completed assessment tool, the present study did at least contribute to this process.

**Future Research**

To further refine the TKIS as a measurement tool, future researchers should address the weaknesses discussed in the limitations section. First of all, extra effort should be made in recruiting a nationally representative sample that spans the full range of superintendent experience, from novices who are preparing to become superintendents to those superintendents who have been in practice for many years. Additionally, this nationally representative sample should consist of those superintendents who are recognized (perhaps at the national level) as being transformational leaders. This will help not only in making findings better able to be generalized, but also give a more accurate measure of the full range of tacit knowledge present in the superintendency as a profession. Furthermore, by utilizing a nationally representative sample of superintendents, items and scenarios will be able to be improved because higher levels of expertise will be able to be sampled. In the present study, only a limited number of
superintendents were available for development of items and scenarios during phase I of the study. Based on the results found in the current study regarding overall distribution and person and item reliability of the items, there is a clear need to gain access to the highest levels of expertise in order to develop items that will span the full range of tacit knowledge in the superintendency.

Additionally, more effort needs to go into the development of scenarios and response options that are considered representative of the types of situations faced by superintendents considered to be the transformational leaders in the field. Experts also need to be identified whose responses can be used to generate the expert profile, which would improve reliability. There also needs to be more effort to generate items that target at different levels of ability, especially those that measure intrapersonal tacit knowledge. As such, future researchers in this area should plan to include at least one additional scenario (with five corresponding response options) per type of tacit knowledge in order to compensate for the inclusion of an additional scenario on the measure for intrapersonal tacit knowledge.

Based on the results of the current study, future researchers also need to consider including interviews with the experts who develop the initial scenarios and response options in order to better develop these items up front. Perhaps having some initial feedback during the development of initial tacit knowledge scenarios and items would have reduced the loss of items at later stages of the study, as well as having improved the item reliability estimates.

Lastly, future research in this area should also consider whether there are differences in acquiring and using tacit knowledge to make decisions based on gender, ethnicity, or the size of the district that a superintendent operates. While there is no empirical evidence to support these notions, it is hypothesized that whether an individual is male or female, or from differing ethnic backgrounds, may impact how they approach and ultimately solve problems they face in their daily roles. It may be that tacit knowledge is used differently by individuals depending upon their gender or ethnicity. Additionally, the size or type of district may also impact decision-making and how tacit knowledge is used by the individual to solve problems. A superintendent in a rural district is more apt to think differently about community relationships than a superintendent in a large, urban district. All of these things may interact with how an individual superintendent is likely to use any of the three types of tacit knowledge to be successful.
Conclusion

This study was an important step in bringing the promising construct of tacit knowledge into the context of the superintendency. However, the present study was simply the logical next step in work previously conducted by Nestor-Baker and Hoy (2001) in this area; that of creating an assessment tool for use with superintendents. Development of the TKIS used an existing theoretical framework primarily based on the work of Sternberg and his colleagues. In addition, the use of the partial-credit Rasch model helped to improve the tacit knowledge items contained on the assessment as measurement tools. The final 33 items demonstrated unidimensionality and also demonstrated at least acceptable levels of a hierarchy in which there was some logical order where certain items were found to be easier than others. Although reliability estimates were low for all three measures, there are steps that can be taken to improve these reliabilities, including the addition of more difficult items, and the use of a more diverse sample in future research.

The findings and methods (specifically the application of the Rasch model) in the current study offer exciting promise about the study of tacit knowledge in the superintendency. The hope is that by developing a valid and reliable measure, researchers can better understand how tacit knowledge develops and exactly how it influences decision-making in school superintendents. Ideally, in understanding the mechanisms involved in the acquisition and use of tacit knowledge, researchers and practitioners can improve the training and development of new superintendents and ultimately improve the future state of education.
References


Appendix A: Final Version of TKIS (33 Items)
<table>
<thead>
<tr>
<th>Scenario 1</th>
<th>1. A new board of education member is elected with a personal agenda. To...</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2. The minority has the opportunity to make decisions with the support of...</td>
</tr>
<tr>
<td></td>
<td>3. The minority is feeling concerned about the new agenda. Their...</td>
</tr>
<tr>
<td></td>
<td>4. The majority votes to increase the budget for the arts. The minority...</td>
</tr>
<tr>
<td></td>
<td>5. The minority votes to cut funding for the arts. The majority...</td>
</tr>
<tr>
<td></td>
<td>6. The executive director votes to increase the budget for the arts. The...</td>
</tr>
</tbody>
</table>

**Response Options**

- 1. Strongly Agree
- 2. Agree
- 3. Neutral
- 4. Disagree
- 5. Strongly Disagree

**TACIT Knowledge**

- Tacit = Internal
- Explicit = External

**Sections**

1. Executive Director
2. General Member
3. Smaller Member
4. Majority Member
5. Minority Member
6. Executive Director
Tacit Knowledge

Scenario 2

You notice a child from the elementary school playing in your backyard.

Response Options

1. You never notice the teacher in the other classroom area for the next few years.
2. You observe the teacher in the other classroom area for the next few years.
3. The teacher is a computer teacher who also teaches computer science.
4. The teacher is a math teacher who also teaches math.
5. The teacher is a reading teacher who also teaches reading.
6. The teacher is a science teacher who also teaches science.
7. The teacher is a physical education teacher who also teaches physical education.

You meet with the teacher and announce that he is the head of the math department.

The teacher is a high school teacher and has been teaching for 20 years.

The teacher is a middle school teacher and has been teaching for 10 years.

Please rate the effectiveness of each method of finding the situation based on the following scale:

1 = Extremely ineffective
2 = Ineffective
3 = Somewhat ineffective
4 = Neutral or effective
5 = Somewhat effective
6 = Extremely effective
7 = Extremely effective
Tacit Knowledge

Please rate the effectiveness of each way of handling this situation based on the following scale:

1. Extremely ineffective
2. Very ineffective
3. Ineffective
4. Satisfactory
5. Very effective
6. Extremely effective

Response Options

Following month:

[Given the context, the response options are not fully visible or legible in the image.]
Please rate the effectiveness of each way of handling the situation based on the following criteria:

1. Extremely ineffective
2. Fairly ineffective
3. Neutral
4. Fairly effective
5. Extremely effective

Situation: The budget for the school is tight due to low enrolment. The principal wants to cut the band program.

Scenario 1: The students protest and demand the band program be kept.

Response Options:

1. You notified the YMCA and the schools meet in order to get
   continued in the program.
2. The principal announces the program is closed effective immediately.
3. You check the budget and the schools meet in order to get
   continued in the program.
4. You check the budget and the schools meet in order to get
   continued in the program. The YMCA will only pay for 4 of the 5 grades.
5. You check the budget and the schools meet in order to get
   continued in the program. The YMCA will only pay for 4 of the 5 grades.
Tacit Knowledge

1. Tacit Knowledge

2. Tacit Knowledge

3. Tacit Knowledge

4. Tacit Knowledge

5. Tacit Knowledge

6. Tacit Knowledge

7. Tacit Knowledge

8. Tacit Knowledge

9. Tacit Knowledge

10. Tacit Knowledge

11. Tacit Knowledge

12. Tacit Knowledge

13. Tacit Knowledge

14. Tacit Knowledge

15. Tacit Knowledge

16. Tacit Knowledge

17. Tacit Knowledge

18. Tacit Knowledge

19. Tacit Knowledge

20. Tacit Knowledge

21. Tacit Knowledge

22. Tacit Knowledge

23. Tacit Knowledge

24. Tacit Knowledge

25. Tacit Knowledge

26. Tacit Knowledge

27. Tacit Knowledge

28. Tacit Knowledge

29. Tacit Knowledge

30. Tacit Knowledge

31. Tacit Knowledge

32. Tacit Knowledge

33. Tacit Knowledge

34. Tacit Knowledge

35. Tacit Knowledge

36. Tacit Knowledge

37. Tacit Knowledge

38. Tacit Knowledge

39. Tacit Knowledge

40. Tacit Knowledge

41. Tacit Knowledge

42. Tacit Knowledge

43. Tacit Knowledge

44. Tacit Knowledge

45. Tacit Knowledge

46. Tacit Knowledge

47. Tacit Knowledge

48. Tacit Knowledge

49. Tacit Knowledge

50. Tacit Knowledge

51. Tacit Knowledge

52. Tacit Knowledge

53. Tacit Knowledge

54. Tacit Knowledge

55. Tacit Knowledge

56. Tacit Knowledge

57. Tacit Knowledge

58. Tacit Knowledge

59. Tacit Knowledge

60. Tacit Knowledge

61. Tacit Knowledge

62. Tacit Knowledge

63. Tacit Knowledge

64. Tacit Knowledge

65. Tacit Knowledge

66. Tacit Knowledge

67. Tacit Knowledge

68. Tacit Knowledge

69. Tacit Knowledge

70. Tacit Knowledge

71. Tacit Knowledge

72. Tacit Knowledge

73. Tacit Knowledge

74. Tacit Knowledge

75. Tacit Knowledge

76. Tacit Knowledge

77. Tacit Knowledge

78. Tacit Knowledge

79. Tacit Knowledge

80. Tacit Knowledge

81. Tacit Knowledge

82. Tacit Knowledge

83. Tacit Knowledge

84. Tacit Knowledge

85. Tacit Knowledge

86. Tacit Knowledge

87. Tacit Knowledge

88. Tacit Knowledge

89. Tacit Knowledge

90. Tacit Knowledge

91. Tacit Knowledge

92. Tacit Knowledge

93. Tacit Knowledge

94. Tacit Knowledge

95. Tacit Knowledge

96. Tacit Knowledge

97. Tacit Knowledge

98. Tacit Knowledge

99. Tacit Knowledge

100. Tacit Knowledge
You decide to deploy an evaluation system of the Tacit Knowledge document you are about to develop and the following options for each document you are going to develop with their levels of confidence and expertise.

**Response Options**

1. Extensive Review
2. Thorough Review
3. Somewhat Review
4. Need for Improvement
5. No Significant Improvement
6. Scenario does not apply

Please rate the effectiveness of each way of finding this solution based on the following 1 to 7-point scale:

1. Extensive Review
2. Thorough Review
3. Somewhat Review
4. Need for Improvement
5. No Significant Improvement
6. Scenario does not apply

Please rate the effectiveness of each way of finding this solution based on the following 1 to 7-point scale:

1. Extensive Review
2. Thorough Review
3. Somewhat Review
4. Need for Improvement
5. No Significant Improvement
6. Scenario does not apply

Please rate the effectiveness of each way of finding this solution based on the following 1 to 7-point scale:

1. Extensive Review
2. Thorough Review
3. Somewhat Review
4. Need for Improvement
5. No Significant Improvement
6. Scenario does not apply

Please rate the effectiveness of each way of finding this solution based on the following 1 to 7-point scale:

1. Extensive Review
2. Thorough Review
3. Somewhat Review
4. Need for Improvement
5. No Significant Improvement
6. Scenario does not apply

Please rate the effectiveness of each way of finding this solution based on the following 1 to 7-point scale:

1. Extensive Review
2. Thorough Review
3. Somewhat Review
4. Need for Improvement
5. No Significant Improvement
6. Scenario does not apply

Please rate the effectiveness of each way of finding this solution based on the following 1 to 7-point scale:

1. Extensive Review
2. Thorough Review
3. Somewhat Review
4. Need for Improvement
5. No Significant Improvement
6. Scenario does not apply
Tacit Knowledge

---

END - THANK YOU AGAIN FOR YOUR PARTICIPATION ---

Suggestion for change:

Talk with the principal about the issues and make

Turn the agenda over to the Board of Education

Demonstration #2 of Trust

After considering the solution, you decide that the most

Response Options

Please rate the effectiveness of each way of handling the situation based on the

Following 1 to 7 - not effective

1. 7 = Extremely effective

2. 6 = Effectively

3. 5 = Somewhat effective

4. 4 = Neutral or ineffective

5. 3 = Somewhat ineffective

6. 2 = Ineffective

7. 1 = Extremely ineffective

---

You are a superintendent of a high performing school district. All district staff

Senten 6