



**ANALYSIS OF THE COACHING KNOWLEDGE SURVEY
USING THE 7-POINT SCALE**

EVALUATION REPORT

PREPARED FOR:
MONTANA STATE UNIVERSITY
DEPARTMENT OF MATHEMATICAL SCIENCES
WILSON 2-299B
BOZEMAN, MT 59717-2400

SEPTEMBER 2014



ANALYSIS OF THE COACHING KNOWLEDGE SURVEY USING THE 7-POINT SCALE

(JESSE, SUTTON, AND SHTIVELBAND, 2014)

EXECUTIVE SUMMARY

This report describes a number of approaches that were taken in 2013-2014 to validate the Coaching Knowledge Survey (CKS), and extends earlier work done by Jesse, Sutton and Linick (2014) that examined dichotomous response options used to convert to 7-point scale item responses. Expert review of items resulted in dichotomization of items that conformed to the coaching literature versus those that did not. Examining Mathematics Coaching (EMC) Project participants completed multiple instruments several times over the course of the project. Extensive statistical analyses of the dichotomous items suggested negative correlations with outcomes of interest in contrast to what was predicted. Findings also revealed that there might be utility in conducting cognitive interviews with high performing coaches after they completed the items for the last time in the project and revisiting results in light of what was learned in those interviews. The interviews suggested some issues with some items, but, more importantly, indicated that contextual factors—the climate in which coaching occurred—were important when surveys were completed. Issues related to negative correlations remained, as did some ambiguity about what it means to conform to the literature.

Three data sets were utilized in this effort: Sample 1 ($N = 252$)¹ includes a convenience sample of pilot test respondents pooled with respondents that were EMC coaches before the project began; Sample 2 consists of repeated measures of CKS and other coaching effectiveness measures aggregated at the coach level ranging from about 40 to 50 participants; and Sample 3 ($N = 4$) includes results from detailed cognitive interviews conducted with high performing coaches who participated in the project.

The major focus of the validation process described here was to revisit the use of the 7-point scoring option. While there was a logical rationale for dichotomizing items, further exploratory and confirmatory analyses revealed that a 7-point scoring option provided the kind of information needed to continue to move toward adding to the body of evidence needed to establish validity.

Based upon the analyses presented here, it is recommended that the shortened CKS instrument that is included in the appendix should be used as a 7-point scale. It has three scales, and the high level coaching and the low level coaching scales are backed by stronger reliability and validity evidence than the third context of coaching scale that includes measures about working with building principals. The development of this scale should continue, starting with expert review by those who understand how mathematics coaching of teachers actually happens in schools and how principals get involved. This will result in information that can be used to adjust existing items on this scale and to identify new information targets for the potential development of new items.

¹ N is the total number in a sample

RATIONALE

In a previous investigation, EMC explored the following research question: To what extent does a coach's depth of knowledge in two primary domains (coaching knowledge and mathematics content knowledge) influence coaching effectiveness? While the answer was complex, it resulted in a 39 item survey of coaching knowledge grounded in the theory of several prominent researchers (e.g., Knight, 2007; West & Staub, 2003; Costa & Garmston, 2002). Structural Equation Modeling (SEM) results indicated that the relationship between latent coaching effectiveness and the CKS was negative and statistically significant. That is, the CKS measure did not predict coaching effectiveness as expected. In fact, higher scores on the CKS were related to lower coaching effectiveness. Therefore, the structure of the CKS was explored in some detail to discover the nature of this unpredicted relationship.

In addition to the CKS, a number of other instruments were administered to coaches and the teachers they coached over a five-year period. Teacher survey (TS) measures were completed throughout the project. Teachers were also observed by trained data collectors using structured protocols. These measures have all been linked in a longitudinal data set, with teachers nested within coaches, and coaches nested within training cohorts. Coaches were randomly assigned to training cohorts.

As noted, Jesse (2013) and Greenwood (2013) found specific negative relationships between the CKS and other measures of coaching effectiveness. There was a negative and significant relationship between the CKS and a latent coaching effectiveness measure that consisted of coach knowledge of content, coaching behaviors, teacher perceptions, knowledge, beliefs, and teacher behaviors as documented by formal observations using a structured protocol. Greenwood's (2013) investigation showed a negative relationship between raw and decomposed CKS scores and teacher Mathematical Knowledge for Teaching (MKT) scores, while scores centered at the coach level across time indicted a modest positive relationship. These explorations were only partial considerations and do not address the effects after accounting for other variables in the model. It was further reported that coach-level average CKS values had a negative relationship to teacher MKT responses in the same model, and that higher coach average CKS scores were related to lower teacher MKT scores whereas the time-varying CKS scores are estimated to show increases in the teacher MKT scores. This suggested that increases in CKS over time were related to increases in teacher MKT scores with lower CKS scoring coaches being related to lower MKT scoring teachers.

This paper describes ongoing efforts to partially answer components of two primary research questions and several secondary research questions. Accumulating evidence for the validity of the CKS is also described. The primary research questions addressed are:

RQ1: To what extent does a coach's depth of knowledge in two primary domains (coaching knowledge and mathematics content knowledge) influence coaching effectiveness?

RQ3: To what extent are the effects of targeted professional development on coaching effectiveness explained by increases in coaching knowledge and mathematics content knowledge?

Additionally, there was interest in identifying coaches who had teachers with particularly high growth on outcome measures across the project to anchor other efforts to produce validation evidence for the

CKS. Efforts to do this are described in Jesse, Sutton and Linick, 2014. This paper describes continuing efforts to link coaching knowledge to two measures of coaching effectiveness at the teacher outcome level: perceptual teacher survey data and teacher behaviors documented during formal observations conducted by trained observers. Secondary research questions, derived from the primary research questions, include the following:

- How are items on the CKS related to teacher perceptions?
- Which items on the CKS predict which structured professional development experiences provided by the project coaches have had?
- Which items on the original CKS can be used to constitute a one-dimensional scale for measuring coaching effectiveness?
- If a one-dimensional scale for measuring coaching effectiveness was not sufficient, what is the appropriate number of dimensions and how should they be constructed?
- What reliability evidence exists for the proposed revision of the CKS?
- What evidence exists that the revised CKS demonstrates predictive, convergent, and concurrent validity?

METHODOLOGY

In an effort to establish the validity of the CKS, predictive, convergent, and concurrent validity were considered in some detail. Following is a brief description of each.

Predictive Validity. In theory, the CKS should be a predictor of coaching effectiveness. High correlations between CKS and other later measures of coaching effectiveness would provide evidence that the measure has predictive validity. The EMC data set affords a unique opportunity to calculate these correlations with multiple scales, subscales, and individual items measured after initial CKS scores were obtained. Such an analysis framework would identify the ability of the CKS to predict later scores of similar measures of coaching effectiveness.

Convergent Validity. Many of the EMC measures have been collected at the same point in time, or relatively close points in time. This affords the opportunity to determine whether and how strong the relationship is between the CKS and other measures of coaching effectiveness at the same time. That is, teacher survey results should be positively correlated to the CKS, and classroom observation results should also be correlated with the CKS if it is a valid measure.

Concurrent Validity. At the beginning of the study from which the second data set was developed, coaches were randomly assigned to one of two cohorts. For the most part, these two cohorts remained intact for five years. After baseline data were collected, one cohort was trained in coaching techniques while the other cohort served as a control. Then the second cohort received content training, while the first cohort served as a comparison group. Eventually both cohorts received all training. Evidence for concurrent validity can be collected by determining the extent to which the CKS distinguishes between the two groups. This was accomplished through discriminant analyses. Another form of concurrent validity can be established by determining which items predict professional development (PD) group membership.

MEASURES USED

Coaching Knowledge Survey (CKS). To measure coaching knowledge, a 40-item CKS, (later reduced to 39 items) grounded in the theoretical research was created. Two different scoring versions of this survey exist: a 7-point scale version, and a version in which 7-point scale items were converted to a “conforming” metric. A value of “0” meant the item response “did not conform” to the coaching literature base, and a “1” indicated that the item response did conform to the literature. A “percent conforming” measure was created for individual conformity of answers to theoretical positions about coaching. The conforming metric was examined in great depth by Greenwood (2014) and by Jesse, Sutton and Linick (2014). The 7-point measure was used in this present study. A copy of the CKS in its entirety is included in the appendix.

Coaching Skills Inventory (CSI). The CSI, originally developed by Yopp (2008) and modified for EMC, is intended to measure a coach’s perspective on her or his own level of effectiveness or confidence with various coaching responsibilities. The data produced from the instrument are reliable and valid (Yopp et al., 2010). To measure coaching skills, a 24-item survey using a 5 point scale measures teacher reports of their perceptions about coach/teacher relationships, coaching skills, mathematics content, mathematics-specific pedagogy and general pedagogy. A series of other questions elicit information

about educator background and practices, including participation in other mathematics and coaching professional development activities.

Teacher Survey (TS). To measure teacher attitudes and dispositions around a number of constructs, a teacher survey was implemented to those coached by project participants. While the measure is multidimensional, a TS score was collected across the course of the project.

Inside the Classroom Observation Protocol (ITC-COP). The ITC-COP is a widely used instrument suitable for documenting teacher behaviors in mathematics classrooms. It was used in this study to formally observe teachers each year of the project. Observers were trained and re-established validity of observations through follow-up trainings throughout the course of the project.

Mathematical Knowledge for Teaching. All participants are asked to complete the MKT Survey of Content Knowledge for Teaching Mathematics (Hill & Ball, 2004). The instrument is designed to assess each teacher's level of mathematical and pedagogical content knowledge. The instrument has been used extensively in research studies and the data produced have been shown to be reliable and valid.

DATA

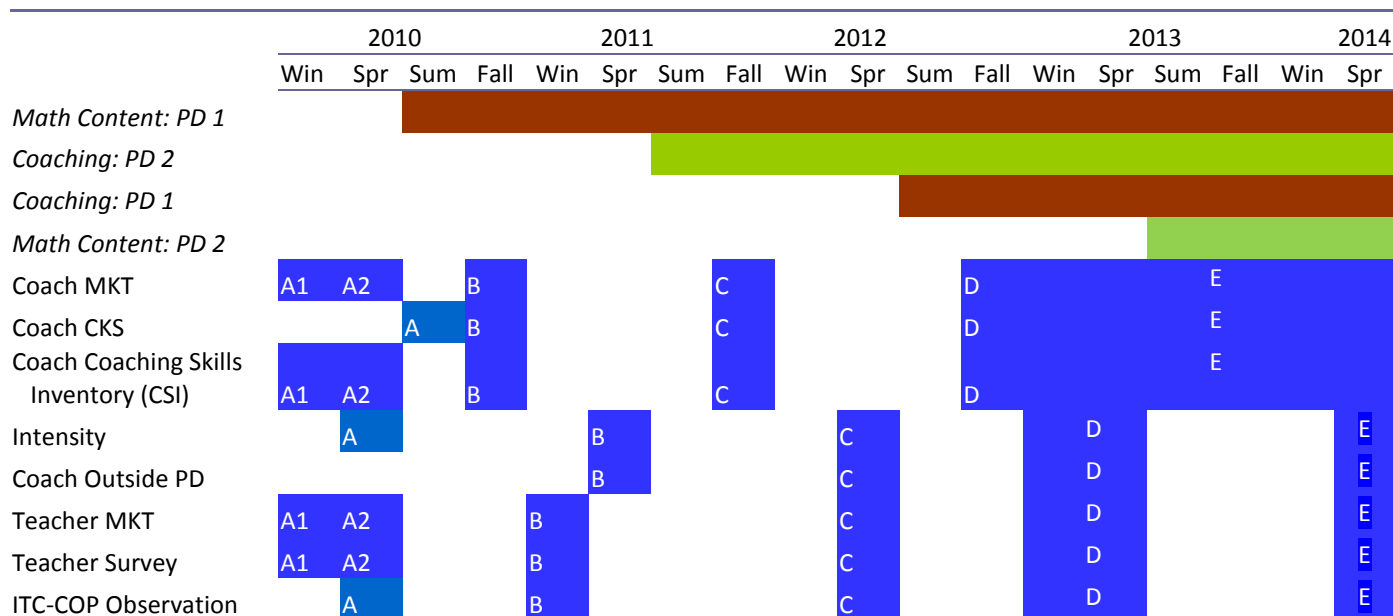
Three different sets of data exist for the validation of the CKS: Sample 1 (item development samples); Sample 2 (coaching study participants), and Sample 3 (cognitive interview participants). Sample 1 was created before the project began and consisted of 252 responses (61 from project coaches before the study, and 191 from a convenience sample of mathematics educators with coaching expertise). Sample 2, which consists of survey and observational data from their teachers over time, utilized almost all of the same measures. Sample 1 was used in Confirmatory Factor Analysis (CFA) and in calculation of reliabilities. Sample 2 was used to calculate relationships with other measures. Sample 3 was selected by identifying coaches who had teachers who demonstrated the most growth on attitudinal and observational measures across the project.

The Sample 2 data collection procedure for this project is complex and proceeded in multiple phases. Exhibit 1 identifies the timeframe for data collection. As noted, pretesting occurred in the winter, spring, and summer of 2010. All of the "A" values represent pretests for teachers and their coaches. The pattern follows with the "B" administrations as the first posttests, the "C" administrations as the second posttests, and the "D" administrations, which were completed in the spring of 2014, as the third posttests. Coaches have also completed the "E" administration of the instrumentation. Teachers have completed six administrations of the teacher survey.

Sample 2 data were collected in the context of a larger study being conducted by EMC. It consists of two different cohorts of coaches, who experience two different interventions in randomized order. The first major intervention occurred in the summer of 2010 for one of two Professional Development (PD) groups. PD Group 1 was a group of coaches randomly assigned to the first cohort. Similarly, PD Group 2 was randomly assigned to the second cohort. Mathematics content PD was provided to PD Group 1 in the summer of 2010, PD Group 2 received coaching PD in the summer of 2011, Group 1 was trained in coaching PD in the summer of 2012, and to complete the cycle of training for Group 2 they were provided PD in mathematics content in the summer of 2013. It follows then, that using A as pretests, B as posttests for PD Group 1, and C as posttest for PD Group 2 is a reasonable approach to address all primary research questions. Using D as a posttest for both groups follows as a next logical step, as did

using E for a posttest for both groups follows as a next logical step to test for PD effects. Complete data sets were obtained from 53 PD Group 1 participants and 34 PD Group 2 participants; although, a number of other coaches participated in the project.

EXHIBIT 1. TIMELINE FOR EMC DATA COLLECTION



Sample 3 data were collected from high performing coaches. By using two growth measures independent of the CKS, coaches who were high performing in comparison to others were identified. A subset of high performing coaches was identified statistically. These growth scores, aggregated at the coach level, were rank-ordered and averaged to create a rank ordering of coaches. The following steps were taken to create the high performing coach subset:

1. We calculated the growth between Time A and Time D on the Teacher Survey for each teacher.
2. We calculated the growth between classroom observation 7-point ratings at Time A and Time D for each teacher.
3. We calculated the mean or average growth on the Teacher Survey and the classroom observation ratings for each coach.
4. We eliminated any coach or teacher who was not consistently paired from the first pretest to the last posttest.
5. We eliminated any coach who did not have data for more than one teacher. All coaches remaining in the sample had pretest and final posttest data for 2 or 3 teachers.
6. We averaged the ranks of Teacher Survey growth and the classroom observation measure growth.
7. We sorted the file by these ranks. There was a natural break in the data for the first six coaches. They were coded as a "1", other coaches were coded as a "0". Four high performing coaches participated in the cognitive interview process.

ANALYSES

Preliminary CFA using all of the 7-point scale items without empirically determined item reversals revealed that the model was not a good fit to the data. This was followed up with a CFA using the polychoric matrix² approach for mixed data sets. Model fit indices, indicator loadings, factor correlations, multidimensional scaling, and reliability analyses were then examined to determine which candidate items should be retained after the trimming process.

Yet, these analyses did not produce a satisfactory one-dimensional solution. Other strategies were then used here with the original 7-point solution to identify suggestions for the optimal number of dimensions, to identify which items may best be used in those dimensions, to identify the direction of coding for all items within the dimensions, to eliminate items for further consideration empirically, to confirm factor structure and to calculate subscale reliability.

First, multidimensional scaling (MDS)³ was utilized with the Sample 1 data set (N= 252) to identify the appropriate number of dimensions. Next, hierarchical cluster analyses⁴ using the furthest neighbor criterion was used to suggest possible subscale structuring. Exploratory factor analysis⁵, set to identify a limited number of dimensions identified by MDS, was then employed. These analyses provided cues for which items should be included in subscales, which should be deleted, and which should be reversed. Confirmatory factor analyses were then conducted to collect evidence about the validity of the subscales, and reliability analyses were used to calculate coefficient alphas⁶ for each subscale.

Utilizing the suggested protocol for describing Structural Equation Modeling (SEM) analyses identified by Brown (2006), Schreiber, et al., (2006) and Cherasaro (2012), the following information was reported: Lisrel Version 8.80 Confirmatory Factor Analysis Results, including Chi Square statistics⁷, Root Mean Square Error of Approximation (RMSEA)⁸, the Comparative Fit Index (CFI)⁹, the Standardized Root mean Square Residual (SRMR)¹⁰, the Non-Normed fit index (NNFI)¹¹, and standardized factor loadings. The *t*

² A polychoric matrix is a correlation matrix used when variables are dichotomous or ordinal and/or involve latent or observed traits.

³ Multidimensional scaling is a statistical technique used for representing the relationships between variables in visual form. It can be used as an alternative to factor analysis to create scales.

⁴ Hierarchical cluster analysis is a statistical technique that groups similar items together to create clusters and organizes these clusters hierarchically.

⁵ Factor analysis is a data reduction technique for clustering multiple variables and creating a smaller group of more reliable factors.

⁶ Coefficient alphas are a measure of internal scale reliability and is a measure of how well multiple items in an instrument measure the same characteristic. Values range between 0 and 1. Items that cluster together well will have a high reliability coefficient (Brown, 2006).

⁷ A Chi-square (χ^2) test is a statistical procedure that is used to examine a test statistic in reference to the chi-square distribution. In this case, the test is used to determine whether Kruskal-Wallis test results were statistically significant.

⁸ RMSEA, or the Root Mean Square Error of Approximation is a statistic for measuring the extent to which a theoretical model fits reasonably well with the population of interest. It is used as fit index for adjusting for model parsimony. Values that are close to 0.06 or below are considered to be a good fit (Brown, 2006).

⁹ In examining baseline comparisons, the CFI depends in large part on the average size of the correlations in the data. If the average correlation between variables is not high, then the CFI will not be very high. A CFI value of .90 or higher is desirable

¹⁰ SRMR, or the Standardized Root Mean Square Residual is a statistic for describing absolute fit of a data set with a theoretical model. It ranges from 0.0 to 1.0, with a value close to 0.08 or below indicating a good fit (Brown, 2006).

¹¹ The NNFI, or the Non-normed fit index, also called the TLI or Tucker Lewis Index, is a statistic for describing comparative fit or incremental fit of a data set with a theoretical model. It has a range that can go beyond the 0.0 to 1.0 range, but is interpreted similarly to some of the other fit indices. A value of .95 or greater indicates a good fit (Brown, 2006).

values and p levels were evaluated but not reported individually since they were all statistically significant.

CKS items were correlated with different groups of criteria: TS scale measures, classroom observations conducted by trained observers, and coach perceptions of skills. Since data were ordinal, and interval in nature, Spearman correlations¹² were calculated. To be conservative, teacher and coach data were aggregated to the coach level to calculate correlations and to conduct discriminant analyses to determine which items predicted professional development group.

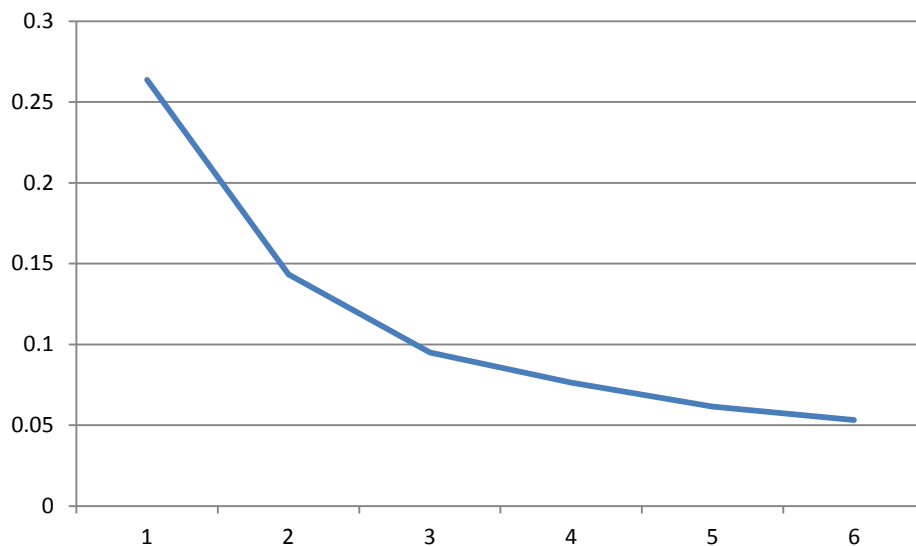
¹² Spearman correlations is a nonparametric measure of statistical dependence between two variables

RESULTS

Study 1: Multidimensional Scaling of Pilot Test Data

Multidimensional scaling is a tool that is useful for identifying gaps in survey constructs, because it can provide visual representations of relationships between items. Items that are close together are similar, and items that are far apart are not similar. In a two-dimensional space, a “circle” of items in this case would be expected if the concept was truly captured. MDS was conducted on the pilot data set with all 49 of the original 7-point response option items by using the ALSCAL routine (a MDS routine) in SPSS version 21. Distances were created from the data via the Euclidian distance option for interval measures, and the level of measurement was specified as interval. The scree plot in Exhibit 2 displays stress values from 1 to 6 dimensions, and suggests that a 2-dimensional or a 3-dimensional solution is best.

EXHIBIT 2. SCREE PLOT OF STRESS BY DIMENSIONS FROM 49 ITEM MDS SOLUTIONS (N=252)



Next, exploratory factor analysis of the 49 items was conducted using varimax rotation¹³ (to force an orthogonal solution) to identify three factors. Exhibit 3 displays the solution. Using the commonly used heuristic of retaining items with loadings of .400 or higher on a factor, results indicate that Factor 1 consists of positive advanced coaching activities, Factor 2 depicts coaching activities that take place with the principal or in the context of the school, and factor 3 describes more rudimentary, even negative coaching behaviors, attitudes and activities. These analyses also suggested that the negative items should be reversed.

¹³ Varimax rotation is a statistical technique used in exploratory factor analyses to make clusters of variables in a scale easier to understand because differences between clusters become more pronounced and easier to identify. Varimax rotation constrains factors to be uncorrelated and is considered to be an orthogonal factor rotation technique (Brown, 2006).

EXHIBIT 3. FACTOR ANALYSIS OF 49 CKS ITEMS ON A 7-POINT SCALE (N = 252)

Question	Item	Factor		
		1	2	3
1a	An effective math coach coaches only on teacher-stated needs	.069	-.335	-.364
1b	Beginning teachers need more coaching than 25 year veterans	.079	-.103	.538
1c	When a teacher says that she or he doesn't want any coaching.	.000	.448	.157
1d	Sometimes an effective math coach has to oppose.	.072	.218	-.134
1e	Teachers will adapt to whatever method of coaching is used.	.058	.045	.451
1f	An effective math coach gets input from a school's principal	-.041	.744	-.084
1g	Number sense is a prerequisite for algebraic thinking	.074	.120	-.143
1h	A coach should put no pressure on teachers to improve their practices	-.034	-.290	-.370
1i	In general, teachers need coaches to model a lesson with a particular strategy	.149	.225	-.372
1j	A teacher can learn new math, but the teacher's basic math intelligence cannot be changed	-.097	.012	.601
2	Once a teacher knows about	-.060	.028	.609
2b	An effective math coach provides teachers	.229	.217	.170
2c	An effective math coach uses state	.180	.483	-.097
2d	An effective math coach asks the principal what	.040	.739	-.074
2e	A student's intelligence can be changed	.189	.044	-.068
2f	Teachers generally have similar teaching styles	.027	-.033	.561
2g	When a teacher says something	.062	-.120	.411
2h	An effective coach sticks to the coaching objectives	.007	.057	.469
2i	Teachers can influence students'	.093	.048	-.277
2j	An effective math coach gives feedback to the principal	-.068	.652	-.280
3a	When a teacher	.402	-.077	-.070
3b	I collect students'	.489	.149	-.043
3c	When decisions about math	.535	.248	.063
3d	I coach teachers on needs	.398	.348	-.063
3e	As a math coach	-.041	-.373	.403
3f	I have difficult conversations	.574	-.069	.034
3g	I always make sure	.530	.050	-.097
3h	I meet with the principal	.533	.453	.052
3i	I encourage teachers to include	.574	.217	-.076
3j	I provide feedback	.465	.519	-.166
4a	I try to provide the teachers	.678	.144	.093
4b	I ask the principal	.220	.720	-.064

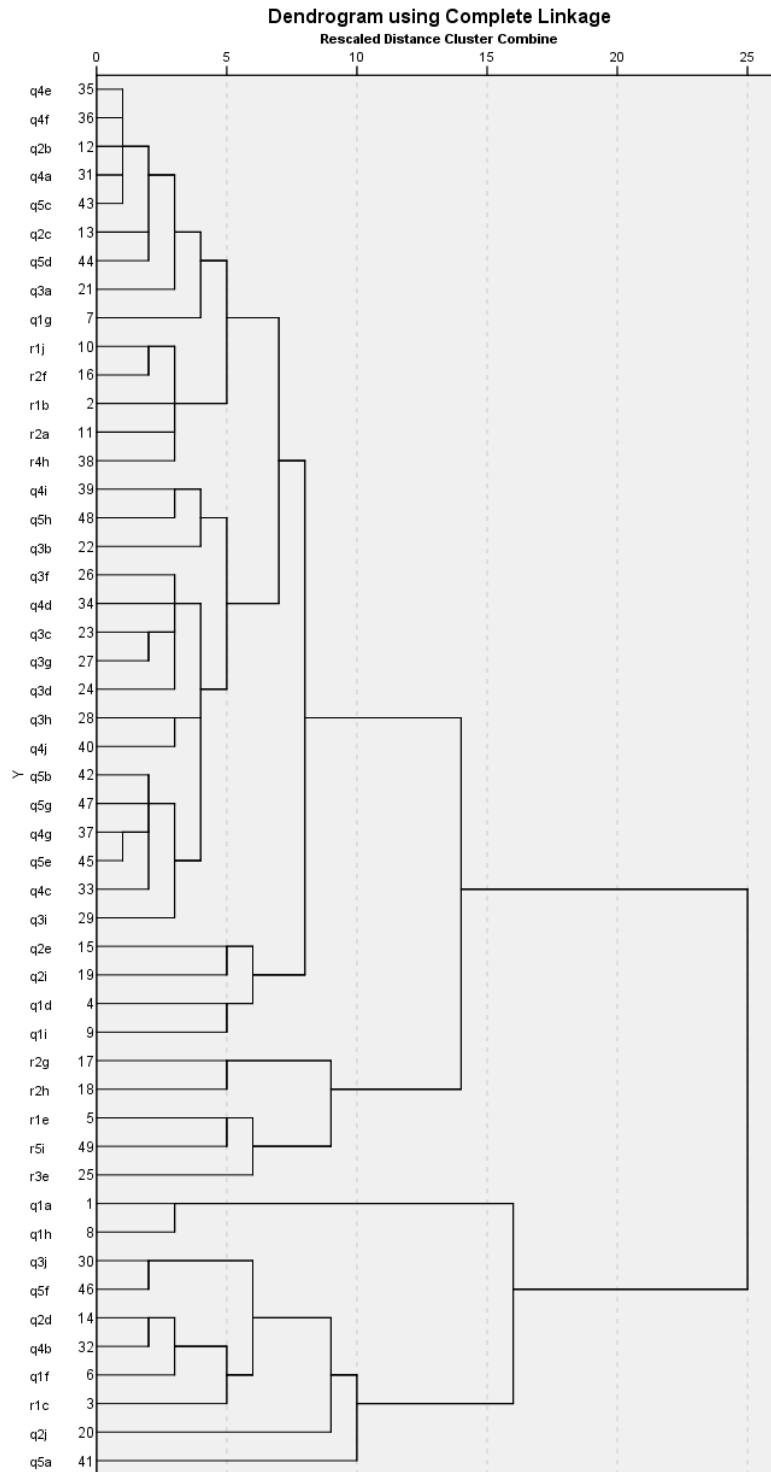
Question	Item	Factor		
		1	2	3
4c	I encourage the teachers	.652	.069	-.162
4d	I help teachers plan	.565	.060	.002
4e	I ask the teachers I coach	.378	.059	-.035
4f	I try to help teachers	.613	.098	.034
4g	I encourage teachers	.695	.057	-.056
4h	I do not alter the coaching plan	.069	-.007	.616
4i	I help teacher identify	.689	.003	-.112
4j	I work with principals	.580	.318	.036
5a	When a teacher says something	.185	.042	-.229
5b	I help teachers reflect	.704	-.030	-.001
5c	I take precautions	.403	.058	.216
5d	I reflect on state assessment	.501	.335	.042
5e	I use student work	.651	.188	.086
5f	I provide feedback	.394	.617	.018
5g	I encourage teachers	.697	-.024	-.030
5h	When a teacher complains	.647	-.020	-.191
5i	I coach my newer teachers more	-.069	-.138	.413

Note. Extraction Method: Principal Component Analysis. Rotation Method: Varimax with Kaiser Normalization. Rotation converged in 5 iterations. Set to 3 factors.

Items with all loadings of less than .400 were discarded from further analyses. Remaining items were used to calculate three subscales: High Level Coaching, Low Level Coaching, and Coaching Context.

Cluster analysis was used as the next multivariate data reduction technique. The dendrogram in Exhibit 4 also suggests three factors in the pilot data set, and also which items might best be used to form these subscales.

EXHIBIT 4. DENDROGRAM OF CKS 7-POINT RESPONSE SCALE ITEMS (N=252)



Note. Item labels on the left refer to questions on the CKS in the appendix. Items with “r” in the label are reversed (see text).

Internal reliabilities as measured by coefficient Alpha were calculated, and all were above the adequate rule of thumb of .70 for the social sciences or psychological tests established by Field (2006) and Kline (1999). Exhibit 5 displays the characteristics of these measures.

EXHIBIT 5. INITIAL MODEL OF COACHING KNOWLEDGE SURVEY TOTAL SCALE (N = 252)

Scale	Items	Mean	SD	Coefficient Alpha
High Level Coaching	q5b, q5g, q4g, q4i, q4c, q5e, q5h, q4f, q4j, q3f, q3i, q4d, q3c, q3g, q5d, q3b, q5c, q3a	5.42	.804	.893
Low Level Coaching	r1b, r1e, r1j, r2a, r2f, r2g, r2h, r3e, r4h, r5i	4.98	.756	.708
Context of Coaching	q3h, q1f, q2d, q4b, q2j, q5f, q3j, q2c	4.80	1.141	.839

Note. Text of items is in the appendix. Low level Coaching items were reversed for all analyses.

Spearman correlations between the three scales were calculated. Exhibit 6 displays the results, and indicates that the correlation between high level coaching and coaching context is highly significant, indicating that the two measures are strongly related. However, the relationship between low level coaching and coaching context is significant and negative, while the relationship between high level coaching and low level coaching is not significant, suggesting that these are two independent measures.

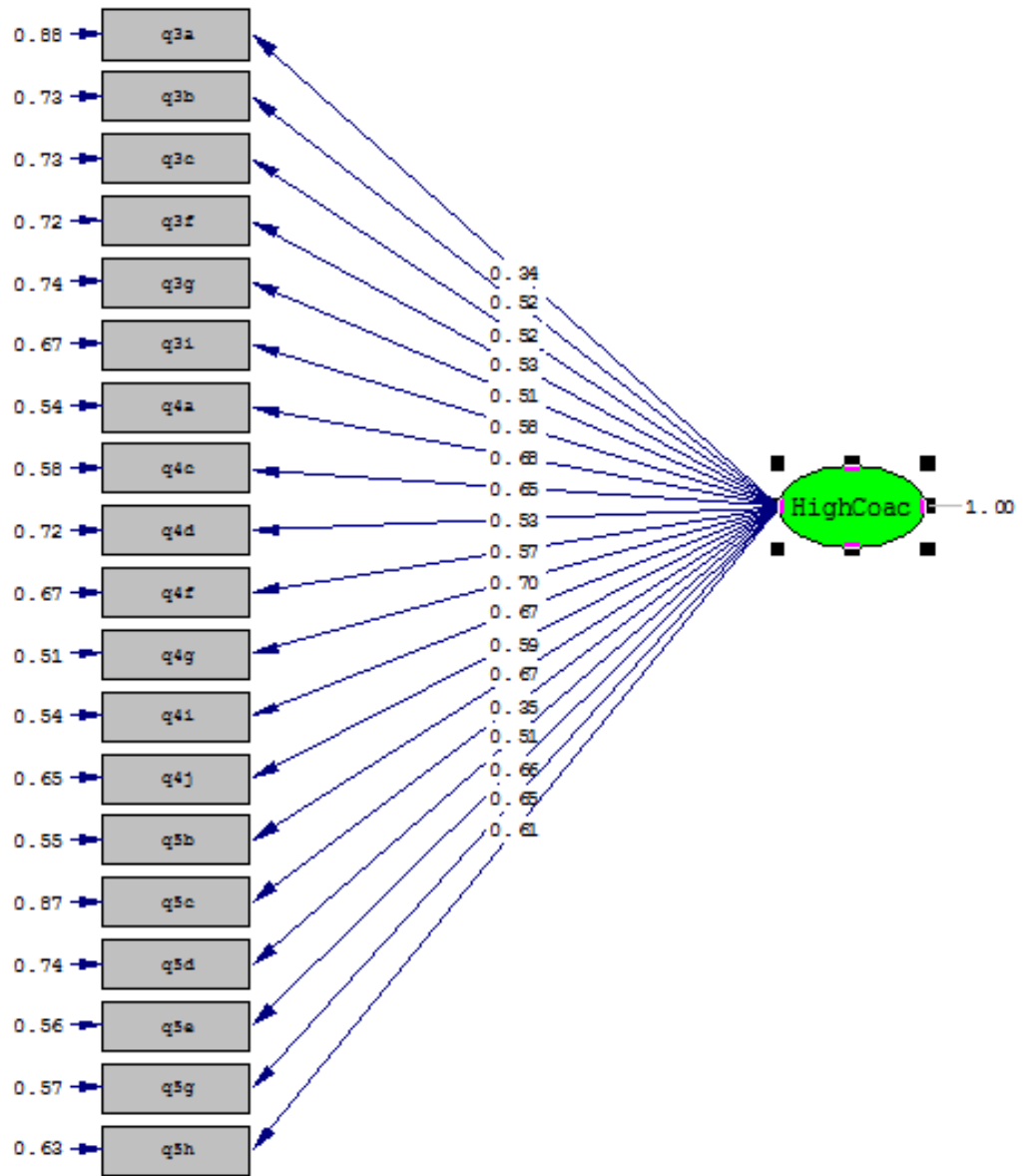
EXHIBIT 6. SPEARMAN CORRELATIONS BETWEEN SCALES (N = 252)

CKS Scale	High Level Coaching	Low Level Coaching	Coaching Context
High Level Coaching	1.000		
Low Level Coaching	.041	1.000	
Coaching Context	.418^{***}	-.181^{**}	1.000

Note: ** $p < .01$, *** $p < .001$

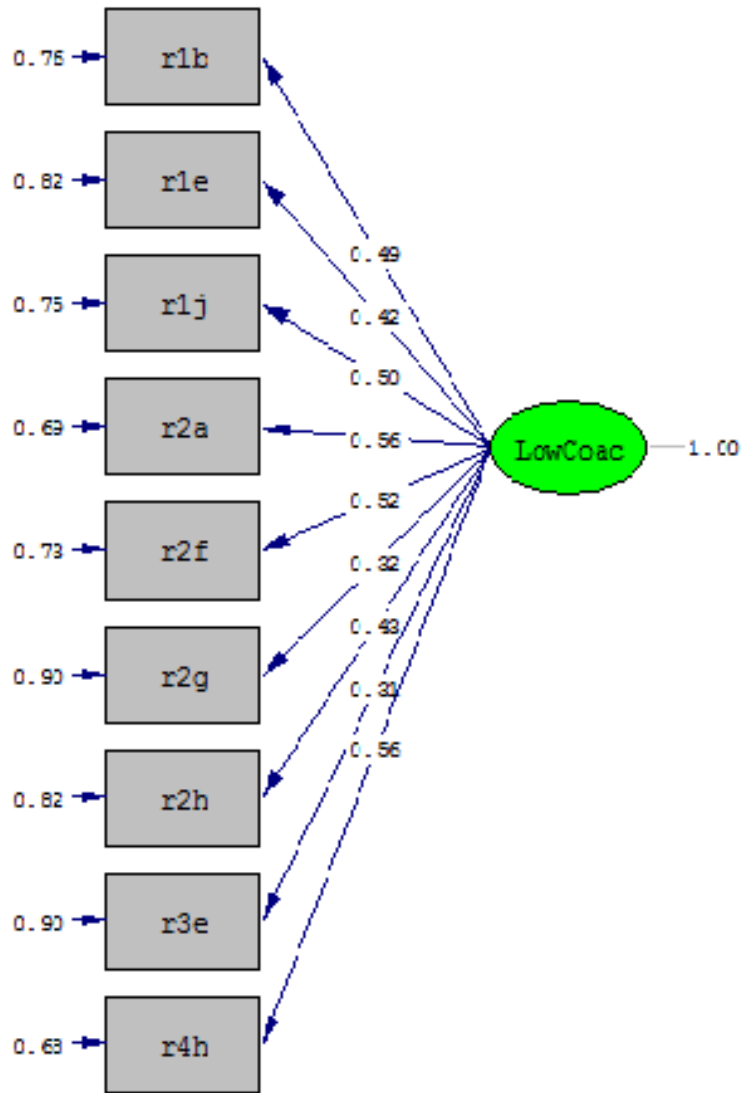
Confirmatory factor analyses were conducted upon the three scales. Exhibit 7 displays findings for the high level coaching scale, and indicates that all items load positively and significantly with the latent construct. Similarly, Exhibit 8 indicates that the loadings for all of the items on the low level coaching scale are positively and significantly related to the latent construct. The items for the context of coaching measure are also all positively and significantly related to this latent construct, as seen in Exhibit 9.

EXHIBIT 7. COACHING KNOWLEDGE SURVEY HIGH LEVEL COACHING SCALE (N = 252)



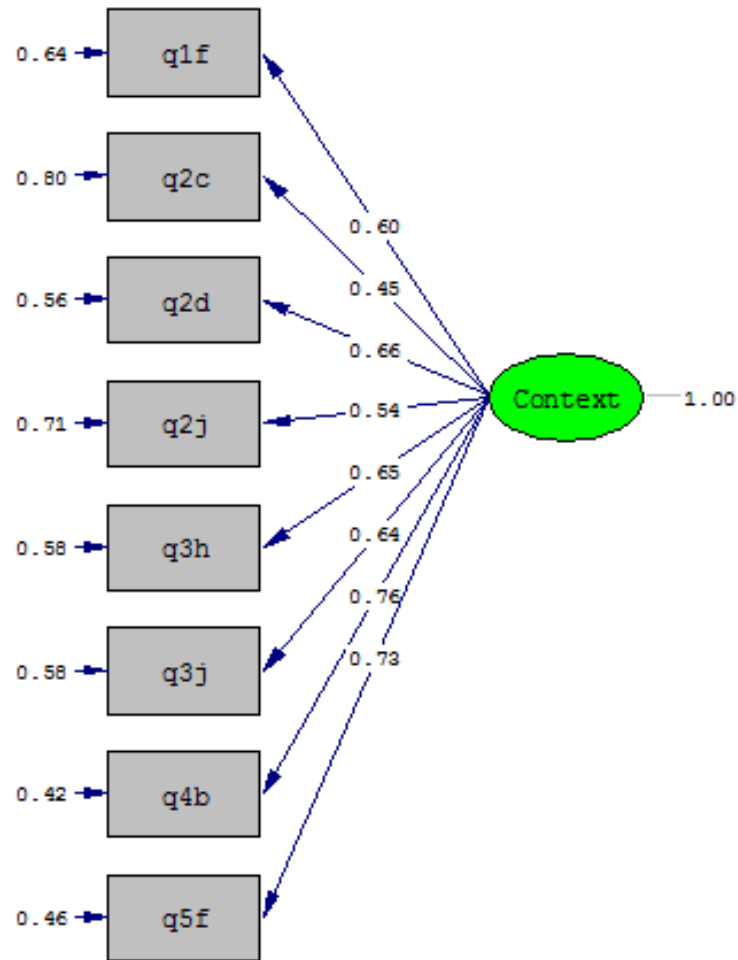
Chi-Square=343.24, df=152, P-value=0.00000, RMSEA=0.071

EXHIBIT 8. COACHING KNOWLEDGE SURVEY LOW LEVEL COACHING SCALE (N = 252)



Chi-Square=56.48, df=27, P-value=0.00075, RMSEA=0.066

EXHIBIT 9. COACHING KNOWLEDGE SURVEY CONTEXT OF COACHING SCALE (N = 252)



Chi-Square=298.65, df=20, P-value=0.00000, RMSEA=0.236

Exhibit 10 displays fit indices for each of the three scales. The high level coaching scale met most of the fit index heuristics, providing evidence that this model was a good fit for the data collected. The low level coaching scale approximated most of the heuristics for fit indices, suggesting the viability of this measure.

EXHIBIT 10. CFA FIT TEST STATISTICS FOR MODELS USING PILOT SAMPLE DATA (N=252)

Fit Index	Heuristic	High level Coaching Scale (18 items)	low Level Coaching Scale (10 items)	Context of Coaching Scale (8 items)
Chi-square	Statistical Significance	343.24***	56.48***	298.65***
Root Mean Square Error of Approximation (RMSEA)	Close to .06 or below	0.071	0.066	0.240
Comparative Fit Index (CFI)	Close to .95 or greater	0.960	0.930	0.820
Standardized Root Mean Square Residual (SRMR)	Close to .08 or below	0.056	0.054	0.100
Nonnormed Fit Index (NNFI)	Close to .95 or greater	0.960	0.900	0.750

Note. *** $p < .001$. Heuristics are from Brown, 2006.

Study 2: ANALYSIS OF PROJECT DATA FOR 5 YEARS

Three different strategies were used to provide additional validity evidence through the use of project data collected longitudinally. As a first step, CKS item Spearman correlations were calculated between items and selected teacher outcomes to provide evidence for predictive and convergent validity. Then, all CKS items were tested to determine whether they could discriminate between coaches trained in Cohort 1 and coaches trained in Cohort 2 to produce concurrent validity evidence.

Results of correlations between teacher survey results and the CKS scale measures are displayed in Exhibit 11. Findings reveal that there were significant correlations for low level coaching measures and teacher survey averages for several measures collected approximately at the same point in time, providing evidence for concurrent validity of this scale. Longitudinal results for the low level coaching scale and the teacher survey scale averages were also significant, providing evidence for predictive validity.

EXHIBIT 11. PEARSON CORRELATIONS BETWEEN TEACHER SURVEY AVERAGES AND CKS SCALES (N = 46)

CKS Scale	Time 1	Time 2	Time 3	Time 4	Time 5	Time 6
High Coaching A	-.048	-.090	-.077	-.094	-.010	-.140
High Coaching B	-.161	-.254	-.137	-.143	-.126	-.244
High Coaching C	-.146	-.264	-.115	-.119	-.025	-.213
High Coaching D	.024	-.108	.031	-.010	.065	-.050
High Coaching E	-.188	-.140	.038	.028	.163	.050
Low Coaching A	.214	.361*	.391**	.268	.437**	.417**
Low Coaching B	.430**	.478**	.489**	.315*	.445**	.399**
Low Coaching C	.426**	.452**	.477**	.368*	.522**	.433**
Low Coaching D	.299*	.336*	.339*	.254	.468**	.358*
Low Coaching E	.284	.287	.228	.125	.369*	.407**
Coaching Context A	-.176	-.244	-.191	-.143	-.215	-.187
Coaching Context B	-.178	-.292*	-.181	-.061	-.164	-.173
Coaching Context C	-.108	-.221	-.123	-.074	-.156	-.158
Coaching Context D	-.072	-.115	-.100	.007	-.130	-.210
Coaching Context E	-.273	-.305*	-.233	-.100	-.087	-.171

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

Exhibit 12 displays Spearman correlations of coach average classroom observation ratings for teachers they worked with and the CKS scale measures. Almost all are not statistically significant, with the exception of the low level coaching measure, which was systematically correlated with several of the classroom observation measures. The absence of these low level coaching perceptions was positively related to classroom observation scores of teachers coached. This was particularly true for the Time 2 low level coaching scores collected after training began, which were predictive of teacher observation average scores at time 2. This constitutes evidence of convergent validity. Time 2 low level coaching scale scores were also correlated with later teacher observation averages. This is evidence of predictive validity.

EXHIBIT 12. SPEARMAN CORRELATIONS BETWEEN CLASSROOM OBSERVATION AVERAGES AND CKS SCALES (N = 46)

CKS Scale	Inside the Classroom Observation Protocol Rating Averages by Coach				
	Rating 1	Rating 2	Rating 3	Rating 4	Rating 5
High Coaching A	.004	-.089	-.185	-.170	.000
High Coaching B	.017	-.363*	-.216	-.282	-.078
High Coaching C	-.011	-.205	-.149	-.189	-.054
High Coaching D	.073	-.237	-.273	-.245	-.222
High Coaching E	.034	-.290	-.131	-.185	-.013
Low Coaching A	.240	-.031	.081	.034	.104
Low Coaching B	.297*	.294*	.285	.325*	.251
Low Coaching C	.372*	.176	.224	.221	.082
Low Coaching D	.343*	.028	.005	-.016	.076

CKS Scale	Inside the Classroom Observation Protocol Rating Averages by Coach				
	Rating 1	Rating 2	Rating 3	Rating 4	Rating 5
Low Coaching E	.180	.219	.197	.261	.357*
Coaching Context A	-.174	.005	-.080	-.240	-.070
Coaching Context B	-.026	-.009	-.173	-.282	-.169
Coaching Context C	-.061	.132	.067	-.077	-.083
Coaching Context D	-.062	.021	-.077	-.139	-.014
Coaching Context E	-.019	.063	-.162	-.241	-.131

Note: * $p < .05$, ** $p < .01$

Finally, discriminant analysis was conducted to determine whether any of the scales (high level coaching, low level coaching and context of coaching) could accurately predict group membership for the coaches in the two cohorts (21 coaches were in each cohort during the fifth year). The process could predict cohort membership with 62 percent accuracy when cross-validated methods were used. (The expected value is 50 percent). However, only one indicator, positive coaching at Time C, was a statistically significant predictor of cohort membership, $F(1, 40) = 7.412, p = .010$. At Time C, PD Group 1 had received math content PD the summer before, while PD Group 2 just finished Coaching Content PD right before the survey.

As a follow-up analysis of concurrent validity, discriminant analysis was performed for each of the five administrations of the CKS by selected items used to create the scales. Exhibit 13 displays the items that significantly distinguished between the two professional development groups. By far, the largest differences between the two groups were at Time C, when there were nine different items that separate the two groups statistically. Seven of these items were on the high level coaching scale. By Time D, both PD groups had Coaching PD. Differences between the two groups dissipated considerably. By time E, both groups had completed the mathematics content and the coaching content PD, and there were only two items that could distinguish the two. These results not only provide additional concurrent validity evidence for selected items in the scales, but also support the notion that the constructs behind the items are malleable and can be influenced by training.

**EXHIBIT 13. DISCRIMINANT ANALYSIS RESULTS BY ITEM FOR COACHES IN PD GROUP 1 (N = 53)
AND PD GROUP 2 (N = 34) BY MEASUREMENT POINT IN TIME**

Item Number	Time A	Time B	Time C	Time D	Time E
	Discriminant Analysis Significance Level				
CKS3 aA	.779	.794	.543	.472	.185
CKS3 bA	.004*	.605	.024*	.366	.448
CKS3 cA	.238	.957	.161	.929	.578
CKS3 fA	.106	.322	.407	.794	.677
CKS3 gA	.543	.487	.644	.207	.817
CKS3 iA	.011*	.535	.892	.953	.901
CKS4 aA	.004*	.220	.227	.057	.624
CKS4 cA	.654	.975	.010*	.366	.243
CKS4 dA	.235	.717	.000***	.004**	.170
CKS4 fA	.832	.947	.003**	.915	.028*
CKS4 gA	.038*	.604	.022*	.208	.826
CKS4 iA	.111	.450	.000***	.118	.933
CKS4 jA	.616	.668	.008**	.109	.785
CKS5 bA	.079	.808	.126	.312	.112
CKS5 cA	.671	.183	.060	.928	.596
CKS5 dA	.623	.918	.401	.567	.524
CKS5 eA	.895	.658	.115	.105	.489
CKS5 gA	.264	.288	.444	.937	.833
CKS5 hA	.837	.485	.096	.979	.144
rCKS1 bA	.476	.179	.120	.793	.177
rCKS1 eA	.819	.146	.015*	.260	.078
rCKS1 jA	.347	.163	.115	.605	.264
rCKS2 aA	.007**	.151	.017*	.773	.451
rCKS2 fA	.808	.162	.531	.168	.291
rCKS2 gA	.340	.024*	.510	.179	.207
rCKS2 hA	.125	.109	.245	.384	.092
rCKS3 eA	.516	.563	.533	.131	.943
rCKS4 hA	.631	.783	.699	.309	.036*
CKS 1fA	.664	.497	.504	.107	.176
CKS2 cA	.272	.299	.750	.729	.745
CKS2 dA	.203	.064	.120	.068	.793
CKS2 jA	.029*	.008*	.883	.123	.056
CKS3 hA	.210	.935	.533	.578	.202
CKS3 jA	.243	.916	.055	.008**	.157
CKS4 bA	.185	.238	.125	.130	.569
CKS5 fA	.197	.844	.070	.008**	.349

Note: * $p < .05$, ** $p < .01$, *** $p < .001$

Study 3: Cognitive Interviews of High Performing Coaches

Based upon previous analyses as described earlier (Jesse, Sutton and Linick, 2014), a set of items that constitute candidates for further scale development has been identified. Exhibit 13 displays the larger pool of these items which were retained for further study and validation.

EXHIBIT 14. CANDIDATE CKS ITEMS FOR VALIDATION IDENTIFIED WITH MULTIPLE TECHNIQUES

Question	Item
1a	An effective mathematics coach coaches only on teacher-stated needs.
1b	Beginning teachers need more coaching than 25-year veterans.
1c	When a teacher says that she or he doesn't want any coaching, an effective mathematics coach respectfully does not try to persuade the teacher to accept coaching.
1d	Sometimes an effective mathematics coach has to oppose school or teacher actions that are not good for students' mathematics learning.
1h	A coach should put no pressure on teachers to improve their practices.
2h	An effective coach sticks to the coaching objectives established with a teacher at the beginning of the year.
3b	I collect students' mathematics work from a teacher's classroom to guide our coaching conversations.
3c	When decisions about mathematics instruction are being made, I ensure that the decision-makers interpret research literature accurately.
3d	I coach teachers on needs that I observe in the teacher, even when the teacher is unaware of these needs.
3f	I have difficult conversations with teachers, when necessary, about mathematics misconceptions they hold.
3h	I meet with the principal to discuss the school's vision for mathematics instruction.
3i	I encourage teachers to include, in each lesson they teach, summaries of what students learned.
3j	I provide feedback to teachers about whether or not the school is meeting its vision for mathematics instruction.
4a	I try to provide the teachers I coach with an understanding of how the mathematics they teach supports learning beyond the grade level they teach.
4c	I encourage the teachers I coach to reflect on similarities and differences among mathematics topics in the curriculum.
4d	I help teachers plan their lessons.
4e	I ask the teachers I coach what aspects of mathematics teaching they need help with
4f	I try to help teachers understand my role as a mathematics coach.

Question	Item
4h	I do not alter the coaching plan developed with the teacher at the beginning of the school year.
4i	I help teachers identify consistencies and inconsistencies between their own practices and the practices recommended by the National Council of Teachers of Mathematics.
5b	I help teachers reflect on discrepancies between espoused beliefs and actual practices.
5c	I take precautions to ensure that my demonstration lessons do not inadvertently send a message that I am the expert and the teacher is not
6	Base 10 Coach Scenario (multiple choice)
10	Teaching Strategy Discussion Scenario (multiple choice)

A convenience sample of four high performing coaches was interviewed to learn more about how they answered the items the way they did, when they did. Data from these interviews were coded using methodologies explicated by Miles, Huberman, and Saldana (2014), and used to make determinations about whether respondents understood the directions, and answered the questions as EMC researchers intended. This information was then used to understand results obtained, and make recommendations for future survey modifications.

The following items were subsequently empirically identified as items to use with a subset of coaches to gain insights into the thinking behind why they responded the way they did in the cognitive interviews based on response patterns and correlations with other measures (see Jesse, Sutton and Linick, 2014):

1. 3d
2. 3h
3. 3j
4. 4b
5. 5f
6. 5h
7. 4i
8. 3b
9. 4d
10. 3i

In order to complete the survey revision cycle, EMC provided the subset of high performing coaches a copy of the survey and asked them to answer the following questions *after* they have completed the last study (modeled in Cherasaro, 2012):

- What problems, if any, did you have completing the survey?
- Are the directions clear? If not, why not?
- Are there any words or language in the survey that coaches might not understand? Please explain.
- Did you find any of the questions redundant or unnecessary? If so, which ones? Why?
- Were any of the questions difficult to answer? If so, why?

- (item by item, or selected/balanced items, 39 items) What did you think this question was asking? How would you phrase it in your own words?
- Do the answer choices allow you to answer as you intended? Please explain.
- Is there anything you would change about the instrument? Please explain.

A total of 10 CKS instrument items were examined using a cognitive interview approach with 4 high performing mathematics coaches who were part of the Examining Mathematics Coaching project. Each item was selected for this analysis due to differences in participant item responses and items were rated on a Likert scale from 1 (not at all descriptive of my coaching practices) to 7 (very descriptive of my coaching practices). As part of the cognitive interview process, each participant reviewed the 10 CKS instrument items and answered the following 4 follow-up questions for each CKS instrument item:

1. What did you think this question was asking? How would you phrase it in your own words?
2. Do the answer choices allow you to answer as you intended? Please explain.
3. What were you thinking about when you responded the way you did to that question?
4. What experience or knowledge influenced the way you responded to that item?

The following section describes themes and patterns that emerged from the cognitive interview qualitative analysis. This section is organized by each CKS instrument item, followed by the overall themes and patterns across items. CKS instrument item findings are listed in the order in which they were asked during the cognitive interviews. In addition, the items ratings are listed following the question as a reference.

CKS INSTRUMENT ITEM QUALITATIVE FINDINGS

3d. I coach teachers on needs that I observe in the teacher, even when the teacher is unaware of these needs. [4, 4, 6, 6].

Overall, this item appeared to be understood by participants. This was evident by the similar ratings ranging from 4 to 6. Differences in item responses appeared for scenarios in which the participant did not have the opportunity to observe teachers as part of their role as a coach. In response to the question *do the answer choices allow you to answer as you intended? Please explain*, one participant shared, “Yes, I knew what was being asked about – but couldn’t answer because of my circumstance.” Another observed difference in interpretation was in response to the question, *what experience or knowledge influenced the way you responded to that item?* Three of four participants referenced coaching; however, one respondent describe how training influenced how they responded. For example, this participant stated the following, “It was the training I had through the project and EDC online course, and my own personal reading of math coaching books – Questioning.”

3h. I meet with the principal to discuss the school’s vision for mathematics instruction. [6, 1, 7, 4]

Overall, this item appeared to be understood by participants. While the ratings ranged from 1 to 7, this range appeared to reflect context rather than different interpretations of the item. For example, in response to the question, *what did you think this question was asking? How would you phrase it in your own words*, the participant with the lowest rating shared that, “In our district, we don’t meet with the principal.” This trend remained in response to the question, *what were you thinking about when you responded the way you did to that question*, in which this same participant shared that, “We have a

confidentiality piece in our program, so we don't meet with our principals." For the question, *what experience or knowledge influenced the way you responded to that item*, one participant responded with, "policy" which was different than other responses that focused on coaching experience? Given the brevity of this response, it is difficult to determine what this respondent meant by policy in response to this question.

3j. I provide feedback to teachers about whether or not the school is meeting its vision for mathematics instruction. [6, 1, 6, 4]

Overall, this item appeared to be understood by participants. While the ratings ranged from 1 to 6, this range appeared to reflect context rather than different interpretations of the item. For instance, in response to the question, *what did you think this question was asking? How would you phrase it in your own words*, the participant described how, "The school as a whole – our focus in on the individual teacher – the school vision is the principal's responsibility." Such a response suggests that this particular participant may not view this item as part of their job as a coach, but that the item itself was understood as intended. A similar trend was observed for the question, *what were you thinking about when you responded the way you did to that question*, in which another participant delineated between their job role and that of the principal. As this participant shared, "Building data is principal data; we [coaches] do classroom data."

4b. I ask the principal what he or she believes the mathematics teachers' need are. [6, 1, 6, 4]

Overall, this item appeared to be understood by participants. While the ratings ranged from 1 to 6, this range appeared to reflect context rather than different interpretations of the item. For example, in response to the question, *what experience or knowledge influenced the way you responded to that item*, it should be noted that all respondents provided different answers. Specifically, participants shared that district committees, district policy, practices, and their coaching role influenced the way they responded to this item.

5f. I provide feedback to the principal about whether or not the school is meeting its vision for mathematics instruction. [3, 1, 5, 5]

Overall, this item appeared to be understood by participants, but may be less relevant for some EMC coaches. While the ratings ranged from 1 to 5, this range appeared to reflect context rather than different interpretations of the item. For example, in response to the question, *what did you think this question was asking? How would you phrase it in your own words*, a couple participants discussed how it was not their role to provide this kind of feedback to the principal. As one participant shared, "My own opinion of how it was going to the principal – that wasn't my role – I was coaching teachers." A similar trend was observed in the question that asked, *what were you thinking about when you responded the way you did to that question*, in that the same two participants emphasized how this CKS instrument item was not part of their job description. For instance, one respondent described that it was, "Not my job to tell the principal if the building is going the right way – not an evaluative role."

5h. When a teacher complains about the school's vision for mathematics, I ask the teacher about her or his vision for mathematics. [6, 6, 4, 5]

This item appeared to be well understood by participants. The ratings were fairly consistent across participants ranging from 4 to 6. One respondent expressed discomfort in the prospect of confronting a teacher who would complain about the school's vision for mathematics. This participant rated this item as a 4 and shared that, "I was thinking about how I also can put up a wall myself with people who don't

agree with our school vision. I can also put up a wall to their beliefs, rather than being a good listener” in response to the question, *What were you thinking about when you responded the way you did to that question?* In the follow-up question, *What experience or knowledge influenced the way you responded to that item?*, this same participant acknowledged that such teachers may need additional coaching to overcome the difficulties they experience.

4i. I help teachers identify consistencies and inconsistencies between their own practices and the practices recommended by the National Council of Teachers of Mathematics. [6, 5, 5, 5]

Overall, this item appeared to be understood by participants. The ratings were fairly consistent across participants ranging from 5 to 6. However, one area in which there were differences in the way that participants responded was the question, *What did you think this question was asking? How would you phrase it in your own words?* A couple participants referred directly to the National Council of Teachers of Mathematics in responding to this question; whereas the other two participants viewed this question more generally. For example, one respondent described, “When you look at NCTM standards – when I am working with a teacher I am helping them see they are doing more direct, less investigative, I look at the big practices of NCTM and how do I help teachers see where they are in relation to these.” Whereas, another participant shared the following statement, “To me, it is the idea that you are reinforcing with feedback has power – promoting good practice consistently, making effort to point that out – identifying those things to work on.”

3b. I collect students’ mathematics work from a teacher’s classroom to guide our coaching conversations. [6, 7, 7, 6]

Overall, this item appeared to be understood by participants, but the context could be clarified. The ratings were fairly consistent across participants, ranging from 6 to 7. However, one area in which there were differences in the way that participants responded was the question, *What did you think this question was asking? How would you phrase it in your own words?* The majority of participants interpreted this question as collecting homework or assignment from the students as an aid toward coaching. However, one participant interpreted this question in the context of conducting a classroom observation to guide their coaching conversations.

4d. I help teachers plan their lessons. [5, 6, 6, 5]

Overall, this question appeared to be understood by participants. The ratings were fairly consistent across participants ranging from 5 to 6. However, one area in which there were differences in the way that participants responded was the question, *What experience or knowledge influenced the way you responded to that item?* Participants discussed a variety of experiences or knowledge that influenced the way they responded to the item ranging from their belief system to their coaching experience. As one participant shared, “Definitely the project workshop, but really, this particular question, I am going back to the online coaching course (EDC) and coaching books I read.”

3i. I encourage teachers to include, in each lesson they teach, summaries of what students learned. [5, 5, 6, 5]

Overall, this question appeared to be less well understood by participants. Although, the ratings were fairly consistent across participants ranging from 5 to 6, the interpretation of this question varied across participants. Specifically, from the question, *What did you think this question was asking? How would you phrase it in your own words,* the majority of participants responded that they would encourage teachers to reflect on what they teach. However, one participant interpreted this question as guiding a

class of students to reflect on what they learned. As this participant shared, “When I look at this, a closing to a lesson, what did we learn today? Do they have time to make sense at the end of the lesson?” Another participant shared that, “It is the idea that we are reflecting on what the kids have learned.” Furthermore, in response to the question, *What were you thinking about when you responded the way you did to that question*, a couple of participants shared that they didn’t have time to encourage this component of coaching; another participant thought about the five practices book, and the last participant shared that, “I knew what we did when we were together, but I wasn’t sure when I left what the teacher did.”

CKS INSTRUMENT FOLLOW-UP QUESTION QUALITATIVE FINDINGS

The next section focuses on themes and patterns that emerged by the four follow-up questions posed in this cognitive interview. Ideas in which to revise the CKS instrument items are discussed.

1. What did you think this question was asking? How would you phrase it in your own words?

For each of the 10 CKS Instrument items, this follow-up question resulted in the most variability between participants. Specifically, items 3b, 3h, 3i, 3j, 4i, and 5f resulted in more than one interpretation or phrasing of the question being asked. This finding may be concerning if it suggests that participants interpreted the meaning of the items differently. This appeared to be the case for items 3b, 3i, and 4i. However, for the remaining items, a difference in this follow-up question reflected context rather than different interpretations of the item itself.

2. Do the answer choices allow you to answer as you intended? Please explain.

Across the 10 CSK Instrument items, respondents reported that the answer choices allowed them to answer as intended. The only exception was found in item 3d in which one participant shared that, “By choosing 4, it was the closest I could get to sometimes I do and sometimes I don’t.” Another observation from this follow-up question is that it may have been more difficult to assign a rating value when the item was not relevant for the participant. For example, as one respondent described, “Again, for me, there is nothing to say I would if there was a vision in place – I was unable to do that for some buildings.”

3. What were you thinking about when you responded the way you did to that question?

This follow-up question produced a variety of responses from participants. Indeed, questions 3h, 3j, 5f, 5h, and 3i each represent items in which not all the participants were thinking about the item in the same way when they responded to the question. For instance, in question 3i, each of the participants was thinking about something different when responding to the question. For the remaining questions, it was typical that only one participant was thinking about something else entirely.

4. What experience or knowledge influenced the way you responded to that item?

Participants shared a wide variety of experiences and knowledge that influenced the way they responded to the item. For instance, in items 3d, 3h, 4b, 4d, and 5h different experience and knowledge influenced the way participants responded to the item. For example, in item 4b, each participant described a different form of experience or knowledge that influenced how they responded to the item. If it is important for the item to draw on specific experiences or knowledge, it may be valuable to consider framing the items in such a way that respondents draw upon similar areas of experience or knowledge in addressing the question.

SUMMARY

Three different data sets were reviewed to extend what was learned about the development and validation of the Coaching Knowledge Survey. The instrument was developed from a theoretical base and reviewed and revised by experts in mathematics coaching. A pilot sample, a sample of coaches in two different cohorts who were trained in coaching and knowledge for mathematics teaching, and a sample of high performing coaches were used in three separate studies to refine and collect validation evidence for the instrumentation. Taken together, these three studies contributed significantly to what is known about coaching, and how instrumentation developed to measure its domains can be refined. The original pilot study had utility for screening items. The second study of two coach cohorts was helpful in determining what factors are malleable in a professional development setting, and also contributed validity evidence. The third cognitive interview study provided valuable guidance for the development of a low level coaching scale, a high level coaching scale, and importantly, a context of coaching scale.

Extensive analyses of pilot study data revealed that a one-dimensional solution to the data was not the best fit. Follow-up analyses revealed that a three-dimensional solution was a more adequate model to explain the data. Reliable scales measuring high level coaching, low level coaching, and the context of coaching were developed from analyses of the pilot data.

This study provides predictive validity evidence for the low level coaching scale (things a good coach does not do) and contributes to what we know about the relationship between coaching knowledge and coaching effectiveness. Low level coaching scale scores are positively and systematically correlated to one of the proxy measures of coaching effectiveness, namely average teacher survey results for coaches which measure teacher perceptions about the experience. These results are relatively consistent across years of the project. There was also a less consistent but positive and significant relationship between the low coaching measure and some of the classroom observation averages calculated for each coach.

CKS items are related to teacher perceptions. Evidence of concurrent validity was found for the low level coaching scale. Positive and statistically significant correlations were found between teacher survey averages and low level coaching scale results at the same points in time throughout the project.

Convergent validity was clearly demonstrated by the ability of the high level coaching scale to discriminate between two randomly assigned professional development groups of coaches at a point in time when one group received coaching training and the other did not. When the second group also received the coaching training, the differences between the two groups dissipated substantially as indicated by the high level coaching scale and seven of its items. This analysis also provided evidence that targeted professional development influenced measures of coaching effectiveness.

Additionally, there was interest in identifying coaches who had teachers with particularly high growth on outcome measures across the project to anchor other efforts to produce validation evidence for the CKS. Efforts to do this are described in Jesse, Sutton and Linick, 2014.

This paper describes continuing efforts to link coaching knowledge to two measures of coaching effectiveness at the teacher outcome level: perceptual teacher survey data and teacher behaviors

documented during formal observations conducted by trained observers. Secondary research questions, derived from the primary research questions, include the following:

The present study strongly suggests the contextual constraints under which coaches operate are influential. Interview data revealed that while high performing coaches knew what to do, they often did not do so because the situations in which they operated did not allow for such practices. School and district policies can often be at odds with what coaches are trying to accomplish, yet high performing coaches find ways to prevail.

Scales created from the 7-point version of the CKS are an improvement over other efforts. They are more reliable and they are positively correlated to some measures, but not others. While some evidence for predictive, concurrent and convergent validity was found, it was not overwhelmingly persuasive and raises new questions. The independence of the CKS scales created for this study is encouraging, but it is not consistent. The fact that low level coaching predicted some outcomes, while high level coaching at Time C predicted cohort membership is curious, and should be explored further.

Some of these findings continue to be at odds with what theory suggests. Empirical determinations of item coding directionality conflicts with what the literature on coaching indicated. Specifically, the analyses described here indicated that three items should have been coded negatively in the original theoretical formulation but were not. Item 1b, Beginning teachers need more coaching than 25-year veterans; Item 1j, A teacher can learn new mathematics, but the teacher's basic mathematical intelligence cannot be changed, and item 2f, Teachers generally have similar teaching styles are all items that should have been coded negatively. These anomalies are worthy of further consideration.

Finally, there is room for additional analyses of the data collected through this project. For example, advanced HLM techniques could be used to isolate variability between and within teacher groups to make better estimates of the relationships between CKS scales and outcomes. Recent work in cross-level measurement invariance between and within groups (Schweig, 2014) informs future efforts to understand how averaging of coach outcomes might be problematic. Variation within teachers and between teacher cohorts under coaches should be more systematically explored to expand what needs to be known about how to effectively measure coaching knowledge.

The present study suggests that high quality mathematics coaching is multidimensional, it is contextual, and it is complex. More needs to be learned in order to help teachers improve.

REFERENCES

- Alteras, V.H., Kostarelis, A., Tsitouridou, M., Niakas, D., & Nicolau, A. (2010). Development and preliminary validation of a questionnaire to measure satisfaction with home care in Greece: An exploratory factor analysis of polychoric correlations. *BMC Health Services Research*, 10: 189. Retrieved from: <http://www.biomedcentral.com/1472-6963/10/189>
- Basto, M., & Pereira, J. M. (2012). An SPSS R-menu for ordinal factor analysis. *Journal of Statistical Software*, 46(4). Retrieved from: <http://www.jstatsoft.org/>
- Bonanomi, A., Ruscone, M. N., & Osmetti, S. A. (n.d.). *The polychoric ordinal alpha, measuring the reliability of a set of polytomous ordinal items*.
- Brown, T. A. (2006). *Confirmatory factor analysis for applied research*. New York: the Guilford Press.
- Byrne, B. M. (1998). *Structural Equation Modeling with LISREL, PRELIS and SIMPLIS: Basic concepts, applications, and programming*. Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.
- Cherasaro, T. (2012). *Examining evaluator feedback in teacher evaluation systems*. Unpublished manuscript, limited availability. Englewood, CO: Marzano Research Laboratory.
- Cordray, D. S. & Pion, G. M. (2006). Treatment strength and integrity: Models and methods. In R. R. Bootzin & P. E., McKnight (Eds.), *Strengthening research methodology: Psychological measurement and evaluation (pp 103-124)*. Washington, DC: American Psychological Association.
- Costa, A. L., & Garmston, R. J. (2002). *Cognitive coaching: A foundation for Renaissance schools*. Norwood, MA: Christopher-Gordon Publishers.
- Field, A. (2006). Reliability analysis
<http://www.sussex.ac.uk/Users/andyf/reliability.pdf>
- Gadermann, A. M., Guhn, M., & Zumbo, B. D. (2012). Estimating ordinal reliability for Likert-type and ordinal item response data: A conceptual, empirical and practice guide. *Practical Assessment, Research & Evaluation*, 17(3). Retrieved from: <http://pareonline.net/getvn.asp?v=17&n=3>
- Greenwood, M., (2013). *Teacher MKT Predictive Model*. Presented at the Research Team Meeting of the Examining Mathematics Coaching Project, Bozeman, MT.
- Greenwood, M. & Jesse, D (2014) *Scoring and then analyzing or analyzing while scoring: An operations of GLMM to an education instrument development and analyses*. Unpublished manuscript.
- Hill, H.C., Schilling, S.G., & Ball, D.L. (2004) Developing measures of teachers' mathematics knowledge for teaching. *Elementary School Journal*, 105, 11-30.
- Hulleman, C. S., & Cordray, D. (2009). Moving from the lab to the field: The role of fidelity and achieved relative intervention strength. *Journal of Research on Intervention Effectiveness*, 2(1), 88-110.
- Jesse, D. (2013, March). *Static and Dynamic Model Analysis of Coaching Effectiveness as a Latent Construct*
- Jesse, D., Sutton, J. & Linick, M. (2014). *Analysis of the Coaching Knowledge Survey: Evidence for Validation and Next Steps*. Unpublished manuscript, Examining Mathematics Coaching Project, RMC Research Corporation, Denver, Colorado.
- Knight, Knight, J. (2007b). *Instructional coaching: A partnership approach to improving instruction*. Thousand Oaks, CA: Corwin Press.
- Kline, R. B. (2011). *Principles and Practice of Structural Equation Modeling*. (3rd ed.). New York: The Guilford Press.
- Kline, P. (1999). *The handbook of psychological testing (2nd Edition)*. London: Routledge.
- MacCallum, R.C., Browne, M. W., & Sugaware, H. M. (1996). Power analysis and determination of sample size for covariance structure modeling. *Psychological Methods*, 1(2), 130-149.

- Miles, M.B., Huberman, A. M., and Saldana, J. (2014). *Qualitative Data Analysis: A methods sourcebook*. 3rd edition. Thousand Oaks, CA: Sage
- Raudenbush, S. W. (2011). *Optimal Design Software for Multi-level and Longitudinal Research* (Version 3.01) [Software]. Retrieved from: www.wtgrantfoundation.org.
- Schreiber, J. B., Nora, A., Stage, F. K., Barlow, E. A. and King, J. (2006) Reporting Structural Equation Modeling and Confirmatory Factor Analysis Results: A Review. *Journal of Educational Research*, 99(6), 323-337.
- Schweig, J. (2014). Cross-level measurement invariance in school and classroom environment surveys: Implications for policy and practice. *Educational Evaluation and Policy Analysis*, 36(3), 259-280.
- West, L., & Staub, F. C. (2003). *Content-focused coaching*. Pittsburgh,
- Wickelmaier, F. (2003). An introduction to MDS. Aalborg, East, Denmark: Aalborg University. Retrieved from: <http://homepages.uni-tuebingen.de/florian.wickelmaier/pubs/Wickelmaier2003SQRU.pdf>
- Yopp, D. (2008). Yopp, D.A. (2008, August). *Strengthening strands: Improving mathematics content and pedagogy in middle school teachers for Minidoka and Cassia Counties evaluation report*. (Unpublished Study)
- Yopp, D. (2010). Yopp, D. A., Rose, J., and Meade, C. (2008). Validity and reliability estimates of a set of mathematics classroom coaching reflection instruments. *Education Evaluation and Policy*, submitted.
- Yopp, D., Burroughs, B., Sutton, J., Swackhamer, L. & Greenwood, M. (2010, December). Construct Reliability and Validity of EMC Instrumentation. Unpublished Manuscript, Unpublished manuscript, Examining Mathematics Coaching Project, Montana State University, Bozeman, MT.
- Zumbo, B. D., Gadermann, A. M., & Zeisser, C. (2007). Ordinal versions of coefficients alpha and theta for Likert rating scales. *Journal of Modern Applied Statistical Methods*, 6, 21-29. Retrieved from: http://educ.ubc.ca/faculty/zumbo/papers/ordinal_alpha_reprint.pdf

APPENDICES

CKS Spearman Correlations with Teacher Survey and ITC COP Observation Results

Exhibit A1: Original CKS Instrument

Exhibit A2: Final CKS Instrument

Exhibit A3: Coach cognitive Interview Instrument

6. Given the following scenario, please select the response that you feel most closely answers the question.

A coach has been working with a teacher on using base-10 pieces for subtraction with regrouping. After several discussions, the coach feels that the teacher still doesn't understand this model and suspects that the teacher needs to take more personal responsibility for conceptual understanding of this topic. In their next meeting, the coach asks the teacher for a verbal explanation of the model. In the explanation, the teacher clearly understands how to remove base-10 pieces as they correspond to subtraction, but the teacher is confused about how to exchange larger place values for smaller ones. Which is the most powerful response to help the teacher take ownership of developing a personal knowledge base?

- a. "You're almost right!" followed by a clear explanation of exchanges.
- b. "Let me paraphrase your explanation," followed by a clear restatement of the teacher's approach.
- c. "It's clear you struggle with some aspects of exchanges. Let's go through this again together."
- d. "You confused me during your explanation about the exchanges. Can you provide a better explanation that helps me clear up that confusion?"

7. Given the following scenario, please select the response that you feel most closely answers the question.

A coach has watched a teacher teach a lesson on ordering fractions. During the short seatwork time, the coach noticed that many of the students were ordering the fractions based on the size of the numerators without considering the denominators. The teacher lectured for most of the class period and did not elicit student comments or examine student thinking. From a coaching perspective, what is the most important consideration?

- a. Making sure that this student misconception is addressed so that students don't leave this class with wrong thinking.
 - b. Making sure that the teacher engages with the coach in a conversation about student thinking and learning.
 - c. Making sure that the teacher uses formative assessment strategies in the next class taught.
 - d. Making sure that the lesson is retaught.
-

8. Given the following scenario, please select the response that you feel most closely answers the question.

A veteran sixth-grade teacher tells an instructional coach that mathematics coaching services are not wanted and definitely not needed. This middle-school teacher of 25 years says retirement is soon approaching and, therefore, changes in instruction “will not occur.” The teacher continues by saying, “I already use the tried-and-true method, which I am more than happy with.” The teacher describes this method as: first, answering homework questions; second, introducing a new topic; third, giving the students some seatwork on the new topic; and fourth, walking around to check the students’ seatwork. The teacher says, “If I notice students struggling with the new topic, I’ll go over it again.” The teacher declares this method to be the “best approach anyone has ever come up with” and feels no need to change anything. “Furthermore,” the teacher says, “I will not attend any more useless professional development sessions.” Based on the information in this scenario, which of the following most likely describes how the teacher affirms beliefs about problem students?

- a. A process in which the teacher acts according to what the teacher believes peers will approve of.
- b. A process in which the teacher steps outside of the situation to reflect on the problem objectively.
- c. A process in which the teacher’s beliefs are determined internally.
- d. A process in which the teacher’s beliefs are substantiated by reflecting on what works.

9. Given the following scenario, please select the response that you feel most closely answers the question.

After meeting for the first time a day earlier, a coach and teacher mutually agreed that the coach can come to the teacher’s classroom to observe a subtraction lesson. While observing the lesson, the coach notices that the students’ understanding of the importance of place value notions in subtraction is weak. The coach also notices that the teacher’s conceptual understanding of the subtraction with regrouping algorithm appears weak. After the lesson, the teacher says to the coach, “Only about half of the students could correctly perform the subtraction procedure. I think the teacher before me didn’t teach subtraction very well.” What should the coach do next?

- a. The coach should postpone the discussion of the teacher’s conceptual understanding because bringing it up this early in their relationship would undermine the trust and rapport in their relationship.
- b. The coach should try to teach the next day’s lesson to the same class, modeling the lesson using base-10 blocks for subtraction.
- c. The coach should explain to the teacher that the students are weak at the subtraction procedure because the teacher didn’t address the conceptual basis of the algorithm in class, and the coach should recommend resources for the teacher to use in class.
- d. The coach should ask if the teacher is more concerned about establishing students’ proficiency with the subtraction algorithm or establishing their conceptual understanding.

Continued on next page...

10. Given the following scenario, please select the response that you feel most closely answers the question.

A coach and teacher have discussed a teaching strategy in detail. The coach feels that the teacher knows enough about the strategy to implement it, and the teacher has developed a plan to implement it. At this point, the coach should:

- a. Develop a plan with the teacher for continued coaching support on the strategy and the possible modeling of the strategy.
- b. Leave the teacher alone to try it out a few times so the teacher can grow comfortable with the strategy and gain ownership of it.
- c. Check on the teacher occasionally to make sure the teacher is using the strategy.
- d. Wait for the teacher to ask for further support to avoid appearing "pushy."

11. Please select the response that you feel most closely answers the question. Which of the following is true about teacher learning?

- a. Teachers come to us with fixed intelligence.
- b. Teacher traits such as intelligence can be influenced by coaches.
- c. Teachers are born with traits such as intelligence that cannot be changed.
- d. Coaches should differentiate coaching based on teacher intelligence quotient assessments.

12. Please select the response that you feel most closely answers the question. Which of the following is true about teachers and professional development without a coaching component?

- a. About 10 percent of teachers remember what is presented in professional development, and 5 percent implement it.
- b. Most teachers remember what is presented in professional development, but only about 5 percent implement it.
- c. About 50 percent of teachers remember what is presented in professional development, and about 50 percent implement it.
- d. Most teachers will try the new teaching strategy once.

You have reached the end of the Coaching Knowledge Survey. If you are finished, please submit your responses according to your instructions.

Burroughs, E., Sutton, J., & Yopp, D. (2010). Coaching Knowledge Survey. Bozeman, MT, and Denver, CO: Examining Mathematics Coaching (Montana State University and RMC Research Corporation). Supported by NSF Discovery Research K 12 Program, Award No.

EXHIBTA-2: FINAL CKS INSTRUMENT

Coaching Knowledge Survey

The following survey is designed to capture information about your beliefs and practices related to instructional coaching.

Enter your name or ID code:

1. Please indicate your level of agreement with each of the following statements, from 1 (strongly disagree) to 7 (strongly agree).

	Strongly Disagree	Disagree	More Disagree than Agree	Neither Disagree nor Agree	More Agree than Disagree	Agree	Strongly Agree
	1	2	3	4	5	6	7
b. Beginning teachers need more coaching than 25-year veterans.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
e. Teachers will adapt to whatever method of coaching is used.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
f. An effective mathematics coach gets input from a school's principal on which teachers need to improve their mathematics instruction.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
j. A teacher can learn new mathematics, but the teacher's basic mathematical intelligence cannot be changed.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Continued on next page...

2. Please indicate your level of agreement with each of the following statements, from 1 (strongly disagree) to 7 (strongly agree).

	Strongly Disagree	Disagree	More Disagree than Agree	Neither Disagree nor Agree	More Agree than Disagree	Agree	Strongly Agree
	1	2	3	4	5	6	7
a. Once a teacher knows about a research-based strategy for improving student learning, the teacher will begin using the strategy.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
c. An effective mathematics coach uses state mathematics assessment data when developing a coaching plan with teachers.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
d. An effective mathematics coach asks the principal what she or he believes the teachers' needs are.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
f. Teachers generally have similar teaching styles.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
g. When a teacher says something that isn't quite mathematically correct, an effective mathematics coach says, "You are almost right," and then gives the teacher a clear explanation of the correct mathematics.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
h. An effective coach sticks to the coaching objectives established with a teacher at the beginning of the school year.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
j. An effective mathematics coach gives feedback to the principal about teachers who are struggling in the classroom.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Continued on next page...

3. Please indicate the degree to which each of the following statements is descriptive of your coaching practices, from 1 (not at all descriptive) to 7 (very descriptive).

	Strongly Disagree	Disagree	More Disagree than Agree	Neither Disagree nor Agree	More Agree than Disagree	Agree	Strongly Agree
	1	2	3	4	5	6	7
a. When a teacher says something I find confusing, I paraphrase what I heard and say it back to her or him.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
b. I collect students' mathematics work from a teacher's classroom to guide our coaching conversations.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
c. When decisions about mathematics instruction are being made, I ensure that the decision-makers interpret research literature accurately.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
e. As a mathematics coach, I support mathematics teachers by tutoring their struggling students.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
f. I have difficult conversations with teachers, when necessary, about mathematics misconceptions they hold.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
g. I always make sure that coaching conversations with mathematics teachers are grounded in the mathematics content.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
h. I meet with the principal to discuss the school's vision for mathematics instruction.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
i. I encourage teachers to include, in each lesson they teach, summaries of what students learned.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
j. I provide feedback to teachers about whether or not the school is meeting its vision for mathematics instruction.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Continued on next page...

4. Please indicate the degree to which each of the following statements is descriptive of your coaching practices, from 1 (not at all descriptive) to 7 (very descriptive).

	Strongly Disagree	Disagree	More Disagree than Agree	Neither Disagree nor Agree	More Agree than Disagree	Agree	Strongly Agree
	1	2	3	4	5	6	7
b. I ask the principal what he or she believes the mathematics teachers' needs are.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
c. I encourage the teachers I coach to reflect on similarities and differences among mathematics topics in the curriculum.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
d. I help teachers plan their lessons.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
e. I ask the teachers I coach what aspects of mathematics teaching they need help with.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
f. I try to help teachers understand my role as mathematics coach.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
g. I encourage teachers to include algebraic thinking in the lessons on number sense and operations.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
h. I do not alter the coaching plan developed with the teacher at the beginning of the school year.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
i. I help teachers identify consistencies and inconsistencies between their own practices and the practices recommended by the National Council of Teachers of Mathematics.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
j. I work with principals or other administrators to form a clear message to teachers about effective mathematics instruction.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Continued on next page...

5. Please indicate the degree to which each of the following statements is descriptive of your coaching practices, from 1 (not at all descriptive) to 7 (very descriptive).

	Strongly Disagree	Disagree	More Disagree than Agree	Neither Disagree nor Agree	More Agree than Disagree	Agree	Strongly Agree
	1	2	3	4	5	6	7
b. I help teachers reflect on discrepancies between espoused beliefs and actual practices.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
c. I take precautions to ensure that my demonstration lessons do not inadvertently send a message that I am the expert and the teacher is not.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
d. I reflect on state assessment data to identify curriculum areas that need to be strengthened.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
e. I use student work when coaching mathematics teachers.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
f. I provide feedback to the principal about whether or not the school is meeting its vision for mathematics instruction.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
g. I encourage teachers to set personal improvement goals for mathematics instruction.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
h. When a teacher complains about the school's vision for mathematics, I ask the teacher about her or his vision for mathematics.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
i. I coach my newer teachers more than experienced ones.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

SCORING KEY:

SCALE	Items
High Level Coaching	q5b, q5g, q4g, q4i, q4c, q5e, q5h, q4f, q4j, q3f, q3i, q4d, q3c, q3g, q5d, q3b, q5c, q3a
Low Level Coaching	r1b, r1e, r1j, r2a, r2f, r2g, r2h, r3e, r4h, r5i (Items reversed)
Context of Coaching	q3h, q1f, q2d, q4b, q2j, q5f, q3j, q2c

Original Survey from Burroughs, E., Sutton, J., & Yopp, D. (2010). Coaching Knowledge Survey. Bozeman, MT, and Denver, CO: Examining Mathematics Coaching (Montana State University and RMC Research Corporation). Supported by NSF Discovery Research K 12 Program, Award No. 0918326.
Modified by Jesse, Sutton and Shtivelband, 2014

Coach Cognitive Interview Protocol

Examining Mathematics Coaching DRK-12

RMC Research Corporation, Denver, CO * Montana State University, Bozeman, MT * University of Idaho, Moscow, ID

Part I. Visitation Information

Date: _____ Observer: _____

Name of Coach: _____

Part II. Coach and Teacher Selection

Selecting and Assigning Teachers to Coaches

- *How were the EMC teachers you coached selected or recruited for the project?*

- *Who selected or recruited the teachers that participated?*

- *What criteria was used to determine which (3) teachers would participate in the EMC project?*

- *Were the 3 participating teachers selected from a pool of interested teachers or were they the first 3 approached? (If they were the first 3 approached, what criteria was used to identify these three teachers should be approached?)*

Part III. Inquiries into Coaching Knowledge Survey

Were any of the questions in the Coaching Knowledge Survey difficult to answer? If so, why?

The following questions had a seven point scale (Strongly disagree-1; Disagree-2; More Disagree than Agree; 4-Neither Disagree nor Agree; 5-More Agree than Disagree; 6 Agree; 7-Strongly Agree).

1c. When a teacher says that she or he doesn't want any coaching, an effective mathematics coach respectfully does not try to persuade the teacher to accept coaching.

- 1) What did you think this question was asking? How would you phrase it in your own words?
 - 2) Do the answer choices allow you to answer as you intended? Please explain.
 - 3) What were you thinking about when you responded the way you did to that question?
 - 4) What experience or knowledge influenced the way you responded to that item?
-

1f. An effective mathematics coach gets input from a school's principal on which teachers need to improve their mathematics instruction..

- 1) What did you think this question was asking? How would you phrase it in your own words?
- 2) Do the answer choices allow you to answer as you intended? Please explain.
- 3) What were you thinking about when you responded the way you did to that question?
- 4) What experience or knowledge influenced the way you responded to that item?

3j. I provide feedback to teachers about whether or not the school is meeting its vision for mathematics instruction.

- 1) What did you think this question was asking? How would you phrase it in your own words?
- 2) Do the answer choices allow you to answer as you intended? Please explain.
- 3) What were you thinking about when you responded the way you did to that question?
- 4) What experience or knowledge influenced the way you responded to that item?

4b. I ask the principal what he or she believes the mathematics teachers' need are.

- 1) What did you think this question was asking? How would you phrase it in your own words?
- 2) Do the answer choices allow you to answer as you intended? Please explain.
- 3) What were you thinking about when you responded the way you did to that question?
- 4) What experience or knowledge influenced the way you responded to that item?

5f. I provide feedback to the principal about whether or not the school is meeting its vision for mathematics instruction.

- 1) What did you think this question was asking? How would you phrase it in your own words?
 - 2) Do the answer choices allow you to answer as you intended? Please explain.
 - 3) What were you thinking about when you responded the way you did to that question?
-

4) What experience or knowledge influenced the way you responded to that item?

9. Given the following scenario, please select the response that you feel most closely answers the question.

A coach has been working with a teacher on using base-10 pieces for subtraction with After meeting for the first time a day earlier, a coach and teacher mutually agreed that the coach can come to the teacher's classroom to observe a subtraction lesson. While observing the lesson, the coach notices that the students' understanding of the importance of place value notions in subtraction is weak. The coach also notices that the teacher's conceptual understanding of the subtraction with regrouping algorithm appears weak. After the lesson, the teacher says to the coach, "Only about half of the students could correctly perform the subtraction procedure. I think the teacher before me didn't teach subtraction very well." What should the coach do next?

- a. The coach should postpone the discussion of the teacher's conceptual understanding because bringing it up this early in their relationship would undermine the trust and rapport in their relationship.
- b. The coach should try to teach the next day's lesson to the same class, modeling the lesson using base-10 blocks for subtraction.
- c. The coach should explain to the teacher that the students are weak at the subtraction procedure because the teacher didn't address the conceptual basis of the algorithm in class, and the coach should recommend resources for the teacher to use in class.
- d. The coach should ask if the teacher is more concerned about establishing students' proficiency with the subtraction algorithm or establishing their conceptual understanding.

1) What did you think this question was asking? How would you phrase it in your own words?

2) Do the answer choices allow you to answer as you intended? Please explain.

3) What were you thinking about when you responded the way you did to that question?

4) What experience or knowledge influenced the way you responded to that item?

If time permits, please ask about the following questions from the CKS.

2h. An effective coach sticks to the coaching objectives established with a teacher at the beginning of the year.

1) What did you think this question was asking? How would you phrase it in your own words?

2) Do the answer choices allow you to answer as you intended? Please explain.

3) What were you thinking about when you responded the way you did to that question?

4) What experience or knowledge influenced the way you responded to that item?

3d. I coach teachers on needs that I observe in the teacher, even when the teacher is unaware of these needs.

1) What did you think this question was asking? How would you phrase it in your own words?

2) Do the answer choices allow you to answer as you intended? Please explain.

3) What were you thinking about when you responded the way you did to that question?

4) What experience or knowledge influenced the way you responded to that item?

3h. I meet with the principal to discuss the school's vision for mathematics instruction.

- 1) What did you think this question was asking? How would you phrase it in your own words?
- 2) Do the answer choices allow you to answer as you intended? Please explain.
- 3) What were you thinking about when you responded the way you did to that question?
- 4) What experience or knowledge influenced the way you responded to that item?

5h. When a teacher complains about the school's vision for mathematics, I ask the teacher about her or his vision for mathematics.

- 1) What did you think this question was asking? How would you phrase it in your own words?
 - 2) Do the answer choices allow you to answer as you intended? Please explain.
 - 3) What were you thinking about when you responded the way you did to that question?
 - 4) What experience or knowledge influenced the way you responded to that item?
-