Evaluation of TEAM Rubric

Report to Tennessee Consortium on Research, Evaluation and Development Vanderbilt University

Catherine A. McClellan
John R. Donoghue

October 2013

VANDERBILT
PEABODY COLLEGE

Tennessee Consortium on Research, Evaluation and Development

Peabody #44 | 230 Appleton Place | Nashville, Tennessee 37203
Phone 615-322-5538 | Fax 615-322-6018 www.tnconsortium.org

TENNESSEE CONSORTIUM
ON RESEARCH, EVALUATION & DEVELOPMENT

VANDERBILT
PEABODY COLLEGE

EXECUTIVE SUMMARY

This paper examines investigation of the TEAM instrument, including analyses of the 2011-2012 data. This report summarizes three sets of activities. The first is a set of analyses evaluating the differences, if any, between scores from announced and unannounced teacher observations. The second set of tasks is geared at recommending the possible simplification of the TEAM rubrics through analyses and expert judgment.

The first activity examined the effect of announced versus unannounced observations, and secondarily, formal versus informal observations. Although several comparisons reached the criterion for statistical significance, the size of the effects was quite small. Based on this data, we conclude that there is little meaningful difference between the announced and unannounced observations.

A secondary observation from the first activity was that the data show evidence of clustering at the level of district. We recommend that, where practical, future analyses should reflect this hierarchical structure in the data.

The second activity was a data mining examination of the 2011-2012 observation scores. For each rubric, internal consistency reliability (Cronbach's alpha) was examined, and exploratory factor analyses were conducted to investigate the underlying dimensionality of the indicator data. The reliability (considering number of indicators) of each of the rubrics was high. Possible reductions in each rubric were identified. However, difference in statistics for the best and worst candidate indicators for removal was small, and our recommendation is that preference be given to theory-based revisions.

The third activity was a theory-based review of the current TEAM instruments. Two content experts reviewed the instrument. Their recommendations were in two sections, general, overarching considerations and detailed suggestions at the rubric-indicator level. These recommendations are presented in the text, and the experts' suggested line-by-line revisions to the instrument are presented in tables in the Appendix.

Overall, we recommend that the TN DOE seriously consider the theory-based recommendations. Investing time in selection of the most highly valued aspects of teaching practice, combined with choosing aligned indicators to create a compact instrument is likely to result in a more effective evaluation system and more informative data. Should the department decide to pursue this avenue, we suggest they consider leveraging the investment of time and effort spent by the content experts who completed this analysis to understand and deeply consider existing TEAM instrument.

INTRODUCTION AND CONTEXT

This report summarizes three sets of activities. The first is a set of analyses, undertaken largely under Activity 1, Topic 2 of Clowder's contract, examined the question of whether scores showed evidence of difference based on whether or not the observation was announced. There is a question of whether knowing that one is about to be observed (Announced visits) might modify teacher behavior, possibly modifying lesson plans or even replacing the original lesson with one that is "more score-able" in an effort to increase scores (alternatively, an anxious professional might do better with an unannounced visit). If subjects of the observation changed their behavior when the observation was announced, scores for announced visits would be expected to differ (presumably be higher) than when the visit was unannounced.

The second and third sets of results are deliverables for Activities 1 and 2 under Topic 1 of Clowder Consulting's contract. These tasks are geared at recommending the possible simplification of the TEAM rubrics. The current TEAM instrument is complex, meant for observation of three types of school personnel, and comprising four rubrics, and possessing multiple indicators within each rubric. Assigning such a large number of scores can impose substantial cognitive load on the observer/rater. As cognitive load increases, observers' ability to reliably assign scores may be compromised. There is concern that amount of material that observers have to consider may cause them to have trouble maintaining the rubric "in their heads", and so their ability to reliably assign scores may be jeopardized.

Activities 1 and 2 were designed to provide complementary views of the current TEAM. Activity 1 can be considered a bottom-up, inductive assessment, and involved analysis of 2011-2012 data.

> The analyses in Activity 1 serve to:
> 1) Describe the overall functioning of the rubrics--means, standard deviations and internal consistency measures of reliability, and
> 2) Examine the relationship between the individual indicator scores, with the goal of recommending some indicators for possible deletion from each rubric.

The analyses examined whether there were indicators on the rubric that contributed little unique information; such indicators would be candidates for elimination. Activity 1 tasks were completely data-driven, seeking relationships in the scores which could be capitalized upon.

The third set of results (from Activity 2) took a theory-driven, deductive approach to simplification of the TEAM rubrics. Two content experts in observation of teaching practice (one an instrument developer specializing in mathematics teaching assessment, and one an elementary school teacher and master-coder for another observation rubric) examined the instrument to make recommendations about re-arrangement and simplification. Task 2 was performed strictly from the viewpoint of theory about pedagogy and design of high-quality scoring rubrics of teaching practice; the content experts did not examine (and indeed, did not have access to) the data examined in Activity 1.

ANNOUNCED/UNANNOUNCED AND FORMAL/INFORMAL OBSERVATIONS

A feature of the TEAM observation data is the combination of announced and unannounced observations. An Announced observation may alter the performance of teachers and behavior of students when compared to an Unannounced observation; it may also influence the observer, as Announced observations may require a pre-conference that can bias the observations.

As a preliminary examination of this effect, simple, paired *t*-tests were conducted. For each teacher, the ratings across indicators from multiple ratings were averaged to form a single Announced, and a single Unannounced observation score for each indicator of each rubric. These mean ratings were then compared using paired *t*-tests. Additional subgroup analysis was conducted by stratifying the paired *t*-tests by type of staff: Apprentice, Professionals, and others (all other types as identified in the data). Note that the Professionalism rubric was excluded in the analysis, as no teacher had both Announced and Unannounced professionalism observations in the data we were provided.

RESULTS

Summary statistics and results of the *t*-test comparisons are presented in Table 1. In terms of numbers of observations, Professional educators yielded the majority of the observations for Instruction/Delivery of Services. The pattern is reversed for scores on Planning and Environment rubrics, where the vast majority of observations were of Apprentices.

There were significant differences in mean observation scores between Announced and Unannounced observations for the General Educators. Announced observations had significantly higher mean ratings in Instruction and Environment (by 0.01 and 0.08 points), respectively. However, given a 5-point scale, these differences may be too small to be substantively meaningful. In contrast, for observations on the Planning rubric, Announced observations had *lower* scores by nearly a quarter point (0.23 points).

For the Instruction/Delivery of Service rubric, the size of the effect[1] is quite small (0.05 for Apprentice and 0.02 for Professional educators). For the Environment rubric, the effect sizes are slightly larger (0.14 for Apprentice and 0.10 for Professional educators). For the Planning, the effect size is slightly larger (0.33 for Apprentice educators, and 0.27 for Professional educators). Ratings for the Library Media Specialist or School Services Personnel did not have Instruction/Delivery of Service scores that differed between Unannounced and Announced.

---

[1] Effect sizes here are computed as the difference in means between the two conditions, divided by the standard deviation of the scores for the Unannounced observations.

Table 1. Comparison of Announced and Unannounced Observations: Paired t-tests by Observation and Evaluation Type

Rubric	Observation	Evaluation	n	Difference		Unannounced		Announced		p-value
Instruction / Delivery of Service	General Educator	Apprentice***	6,620	−0.03	(0.49)	3.40	(0.57)	3.43	(0.58)	< 0.001
		Professional**	38,853	−0.01	(0.54)	3.71	(0.61)	3.72	(0.59)	< 0.001
		Others**	162	−0.09	(0.54)	3.29	(0.78)	3.38	(0.68)	0.003
		Total***	45,635	−0.01	(0.53)	3.66	(0.61)	3.68	(0.60)	< 0.001
	Library Media Specialist	Apprentice*	41	0.13	(0.37)	3.56	(0.55)	3.43	(0.58)	0.026
		Professional	842	0.03	(0.49)	3.86	(0.59)	3.83	(0.59)	0.089
		Others	2	0.38	(0.53)	3.79	(0.77)	3.42	(0.24)	0.500
		Total*	885	0.03	(0.49)	3.85	(0.59)	3.81	(0.59)	0.036
	School Services Personnel	Apprentice	15	−0.06	(0.43)	3.76	(0.55)	3.82	(0.36)	0.590
		Professional	51	−0.05	(0.37)	4.02	(0.60)	4.08	(0.64)	0.329
		Total	66	−0.05	(0.38)	3.96	(0.60)	4.02	(0.60)	0.259
Planning	General Educator	Apprentice***	1,605	0.23	(0.69)	3.63	(0.70)	3.39	(0.72)	< 0.001
		Professional***	501	0.20	(0.73)	3.71	(0.74)	3.51	(0.73)	< 0.001
		Others	14	0.19	(0.61)	3.52	(0.68)	3.33	(0.51)	0.263
		Total***	2,120	0.23	(0.70)	3.65	(0.71)	3.42	(0.72)	< 0.001
	Library Media Specialist	Apprentice	9	0.11	(0.37)	3.93	(0.62)	3.81	(0.71)	0.397
		Professional	15	0.33	(0.60)	4.56	(0.43)	4.22	(0.59)	0.051
		Total*	24	0.25	(0.53)	4.32	(0.59)	4.07	(0.65)	0.031
	School Services Personnel	Apprentice	10	0.13	(0.76)	3.67	(0.59)	3.53	(0.74)	0.591
		Professional	2	0.00	(0.00)	3.83	(1.18)	3.83	(1.18)	NA
		Total	12	0.11	(0.69)	3.69	(0.64)	3.58	(0.77)	0.586
Environment	General Educator	Apprentice***	5,363	−0.10	(0.67)	3.79	(0.73)	3.89	(0.70)	< 0.001
		Professional***	1,441	−0.07	(0.62)	3.86	(0.72)	3.92	(0.68)	< 0.001
		Others	43	−0.06	(0.61)	3.42	(0.93)	3.49	(0.73)	0.494
		Total***	6,847	−0.09	(0.66)	3.80	(0.73)	3.90	(0.70)	< 0.001
	Library Media Specialist	Apprentice	29	−0.14	(0.59)	3.87	(0.69)	4.01	(0.61)	0.220
		Professional	22	0.11	(0.74)	4.31	(0.69)	4.19	(0.72)	0.481
		Total	51	−0.03	(0.67)	4.06	(0.72)	4.09	(0.66)	0.754
	School Services Personnel	Apprentice	14	0.14	(0.45)	4.05	(0.62)	3.91	(0.53)	0.252
		Professional	4	0.06	(0.66)	4.56	(0.24)	4.50	(0.61)	0.861
		Total	18	0.13	(0.48)	4.17	(0.59)	4.04	(0.59)	0.284

Note: *$p<0.05$; **$p<0.01$; ***$p<0.001$. Values represent mean and standard deviation (in parenthesis). "Difference" calculated as (Unannounced − Announced); negative values indicate higher scores for Announced observations.

ACCOUNTING FOR CLUSTERING WITHIN THE DATA

Although direct, the above *t*-test analysis is limited in that it ignores the hierarchical nature of schools; staff members are nested within schools, and schools within district. In general, two educators from the same district are expected to be more similar to each other than two teachers selected at random from the population of teachers in the state. This effect is termed "clustering." Analyses that ignore clustering can yield estimates of variability that are too small, leading to declaring effects "significant" when they are not. Based upon the data, we were able to examine the effect of clustering at the district level. Examining the clustering of mean rating scores due to district revealed noticeable clustering, with intraclass[2] correlation coefficients ranging from 0.06-0.18. Tables A1-A5 in the Appendix give detailed information about the intraclass correlations for the different rubrics.

In addition to Announced/Unannounced, the dataset indicated that observations also differed on "formality." Both Announced and Unannounced observations can also be Formal and Informal (as noted in the data). The combination of Announced/Unannounced and Formal/Informal yields 4 types of observation: Announced formal, Unannounced formal, Announced informal, and Unannounced informal. By incorporating formality of the observation, here, we sought to reduce some of the variability in the system. Table 2 presents basic descriptive statistics for mean rubric rating by type staff member observed and by type of observation, and Table 3 presents statistics for Announced by Formality status of observation.

---

[2] The intraclass correlation is the variance due to group (district) membership divided by the total variance. The statistic can be computed from ANOVA quantities: $\frac{\sigma_\alpha^2}{\sigma_\alpha^2 + \sigma_\epsilon^2}$. As defined here, the value ranges from 0 (no clustering) to 1.0 (all variability in scores is due to group membership).

Table 2. Descriptive Statistics of the TEAM Rubric by Overall Announced, Unannounced, Formal, and Informal Observations

Rubric	Observation	Announced / Formal	n	Mean	SD	Quartiles				
						Min	First	Median	Third	Max
Instruction / Delivery of Service	General Educator	Unannounced	54,181	3.63	0.63	1.00	3.25	3.67	4.08	5.00
		Announced	51,328	3.66	0.61	1.00	3.25	3.67	4.08	5.00
		Informal	5,350	3.64	0.60	1.00	3.25	3.67	4.08	5.00
		Formal	100,159	3.65	0.62	1.00	3.25	3.67	4.08	5.00
	Library Media Specialist	Unannounced	1,064	3.84	0.59	1.08	3.42	3.92	4.25	5.00
		Announced	1,106	3.80	0.59	1.92	3.42	3.83	4.25	5.00
		Informal	124	3.80	0.59	2.17	3.33	3.83	4.25	5.00
		Formal	2,046	3.82	0.59	1.08	3.42	3.83	4.25	5.00
	School Services Personnel	Unannounced	8,080	4.10	0.57	1.18	3.73	4.18	4.55	5.00
		Announced	253	4.10	0.58	2.27	3.73	4.18	4.55	4.91
		Informal	8,263	4.10	0.57	1.18	3.73	4.18	4.55	5.00
		Formal	70	4.02	0.60	2.27	3.73	4.09	4.45	5.00
Planning	General Educator	Unannounced	3,340	3.66	0.71	1.00	3.00	3.67	4.00	5.00
		Announced	54,531	3.63	0.71	1.00	3.00	3.67	4.00	5.00
		Informal	52,169	3.63	0.71	1.00	3.00	3.67	4.00	5.00
		Formal	5,702	3.67	0.73	1.00	3.00	3.67	4.33	5.00
	Library Media Specialist	Unannounced	71	4.16	0.69	2.67	3.67	4.33	4.67	5.00
		Announced	1,072	4.15	0.69	1.67	3.67	4.33	4.67	5.00
		Informal	968	4.14	0.70	1.67	3.67	4.33	4.67	5.00
		Formal	175	4.25	0.65	2.33	4.00	4.33	4.67	5.00
	School Services Personnel	Unannounced	4,561	4.08	0.66	1.33	3.67	4.00	4.67	5.00
		Announced	53	4.00	0.77	2.00	3.33	4.00	4.67	5.00
		Informal	4,592	4.08	0.66	1.33	3.67	4.00	4.67	5.00
		Formal	22	4.06	0.64	3.00	3.67	4.33	4.67	5.00

Table 2. Descriptive Statistics of the TEAM Rubric by Overall Announced, Unannounced, Formal, and Informal Observations (cont.)

Rubric	Observation	Announced	*n*	Mean	SD	Quartiles				
						Min	First	Median	Third	Max
Environment	General Educator	Unannounced	47,252	4.06	0.69	1.00	3.50	4.25	4.50	5.00
		Announced	10,995	3.93	0.70	1.00	3.50	4.00	4.50	5.00
		Informal	45,522	4.03	0.70	1.00	3.50	4.00	4.50	5.00
		Formal	12,725	4.06	0.70	1.00	3.50	4.25	4.75	5.00
	Library Media Specialist	Unannounced	916	4.23	0.60	2.00	3.75	4.25	4.75	5.00
		Announced	190	4.14	0.62	2.00	3.75	4.25	4.75	5.00
		Informal	848	4.22	0.60	2.00	3.75	4.25	4.75	5.00
		Formal	258	4.19	0.63	2.25	3.75	4.25	4.75	5.00
	School Services Personnel	Unannounced	4,522	4.32	0.58	1.50	4.00	4.50	4.75	5.00
		Announced	76	4.11	0.63	2.25	3.75	4.25	4.50	5.00
		Informal	4,534	4.32	0.58	1.50	4.00	4.50	4.75	5.00
		Formal	64	4.25	0.60	2.75	4.00	4.50	4.75	5.00
Professionalism		Unannounced	45,416	4.09	0.63	1.00	3.70	4.10	4.60	5.00
		Announced	7,691	3.96	0.65	1.00	3.50	4.00	4.50	5.00
		Informal	45,416	4.09	0.63	1.00	3.70	4.10	4.60	5.00
		Formal	7,691	3.96	0.65	1.00	3.50	4.00	4.50	5.00

Table 3. Descriptive Statistics of the TEAM Rubric by Announced by Formality Observation

Rubric	Observation	Announced	Formal	n	Mean	SD	Quartiles				
							Min	First	Median	Third	Max
Instruction / Delivery of Service	General Educator	Unannounced	Informal	3,490	3.65	0.62	1.00	3.25	3.67	4.08	5.00
			Formal	50,691	3.63	0.63	1.00	3.25	3.67	4.08	5.00
		Announced	Informal	1,860	3.62	0.57	1.00	3.25	3.58	4.00	5.00
			Formal	49,468	3.67	0.61	1.00	3.25	3.67	4.08	5.00
	Library Media Specialist	Unannounced	Informal	80	3.80	0.62	2.17	3.38	3.83	4.25	5.00
			Formal	984	3.85	0.59	1.08	3.42	3.92	4.25	5.00
		Announced	Informal	44	3.80	0.54	2.75	3.33	3.79	4.25	4.92
			Formal	1,062	3.80	0.59	1.92	3.42	3.83	4.25	5.00
	School Services Personnel	Unannounced	Informal	8,047	4.10	0.57	1.18	3.73	4.18	4.55	5.00
			Formal	33	3.87	0.58	2.73	3.55	3.82	4.27	5.00
		Announced	Informal	216	4.09	0.57	2.36	3.73	4.18	4.55	4.91
			Formal	37	4.16	0.60	2.27	3.82	4.36	4.64	4.91
Planning	General Educator	Unannounced	Informal	2,654	3.67	0.72	1.00	3.33	3.67	4.00	5.00
			Formal	686	3.64	0.69	1.67	3.00	3.67	4.00	5.00
		Announced	Informal	49,515	3.63	0.71	1.00	3.00	3.67	4.00	5.00
			Formal	5,016	3.68	0.73	1.00	3.00	3.67	4.33	5.00
	Library Media Specialist	Unannounced	Informal	60	4.15	0.68	2.67	3.67	4.33	4.67	5.00
			Formal	11	4.24	0.78	2.67	4.00	4.67	4.67	5.00
		Announced	Informal	908	4.14	0.70	1.67	3.67	4.33	4.67	5.00
			Formal	164	4.25	0.65	2.33	4.00	4.33	4.67	5.00
	School Services Personnel	Unannounced	Informal	4,548	4.08	0.66	1.33	3.67	4.00	4.67	5.00
			Formal	13	4.05	0.62	3.00	3.67	4.33	4.33	5.00
		Announced	Informal	44	3.98	0.79	2.00	3.33	4.00	4.67	5.00
			Formal	9	4.07	0.70	3.00	3.67	4.33	4.67	4.67

Table 3. Descriptive Statistics of the TEAM Rubric by Announced by Formality Observation (cont.)
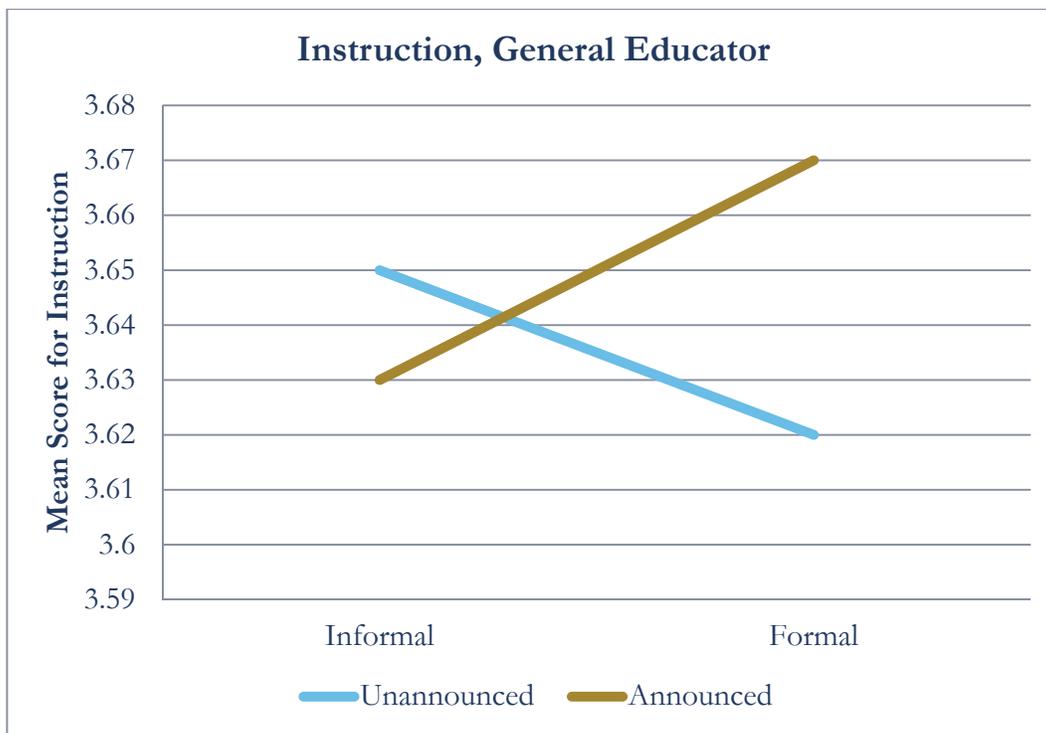
Rubric	Observation	Announced	Formal	*n*	Mean	SD	Quartiles				
							Min	First	Median	Third	Max
Environ-ment	General Educator	Unannounced	Informal	37,343	4.05	0.69	1.00	3.50	4.25	4.50	5.00
			Formal	9,909	4.10	0.69	1.00	3.75	4.25	4.75	5.00
		Announced	Informal	8,179	3.93	0.69	1.00	3.50	4.00	4.50	5.00
			Formal	2,816	3.94	0.72	1.00	3.50	4.00	4.50	5.00
	Library Media Specialist	Unannounced	Informal	727	4.23	0.60	2.00	3.75	4.25	4.75	5.00
			Formal	189	4.21	0.63	2.25	3.75	4.25	4.75	5.00
		Announced	Informal	121	4.13	0.62	2.00	3.75	4.25	4.50	5.00
			Formal	69	4.15	0.62	2.25	3.75	4.25	4.75	5.00
	School Services Personnel	Unannounced	Informal	4,484	4.32	0.58	1.50	4.00	4.50	4.75	5.00
			Formal	38	4.24	0.64	3.00	4.00	4.50	4.75	5.00
		Announced	Informal	50	4.03	0.66	2.25	3.50	4.00	4.75	5.00
			Formal	26	4.26	0.54	2.75	4.00	4.50	4.50	5.00
Professionalism		Unannounced	Informal	45,416	4.09	0.63	1.00	3.70	4.10	4.60	5.00
			Formal								
		Announced	Informal								
			Formal	7,691	3.96	0.65	1.00	3.50	4.00	4.50	5.00

Note: There were no Unannounced/formal or Announced/informal observations for Professionalism.

To account for clustering, we did another set of analyses using a mixed-effects regression. District was treated as a random effect. The fixed effects were Announced status (Announced/Unannounced), Formality (Formal/Informal observation), and the interaction Announced status x Formality. A separate analysis was performed for each type of educator on each of the 3 dimensions. Results are summarized in Table 4.

For General Educator, the interaction of Announced status by Formality interaction was significant for all 3 of the TEAM rubrics. For school services staff, there was a significant main effect for announced status, and a significant announced status by formality interaction. Figures 1-4 present plots of the means for these 4 conditions. Please note that the vertical scale in each case is optimized for the specific data in the figure, and changes for each.

## Figure 1. Mean Instruction Rubric Score by Announced Status and Formality for General Educators

## Table 4. Comparison of Announced and Formal Observations Accounting for District Effects: Mixed-effects regression

| Rubric | Observation | Fixed Effect | | | | | | | | | Random Effect | | | |
| | | Formal vs. Informal | | | Announced vs. Unannounced | | | Interaction | | | Var(Int.) | | Var (Resid.) | |
		Est.	SE	*p*	Est.	SE	*P*	Est.	SE	*p*	Est.	SE	Est.	SE
Instruction / Delivery of Service	General Educator	0.00	(0.01)	0.973	0.00	(0.00)	0.745	0.02	(0.00)	0.000	0.03	(0.00)	0.36	(0.00)
	Library Media Specialist	0.00	(0.04)	0.926	−0.03	(0.03)	0.359	0.00	(0.03)	0.901	0.04	(0.01)	0.31	(0.01)
	School Services Personnel	0.01	(0.04)	0.735	0.01	(0.04)	0.692	0.00	(0.04)	0.943	0.04	(0.01)	0.29	(0.00)
Planning	General Educator	0.01	(0.01)	0.395	−0.01	(0.01)	0.159	0.02	(0.01)	0.013	0.07	(0.01)	0.46	(0.00)
	Library Media Specialist	0.05	(0.06)	0.346	0.01	(0.06)	0.898	0.01	(0.06)	0.831	0.03	(0.01)	0.45	(0.02)
	School Services Personnel	0.08	(0.08)	0.284	−0.01	(0.08)	0.929	0.00	(0.07)	0.973	0.06	(0.01)	0.39	(0.01)
Environment	General Educator	0.02	(0.00)	0.000	−0.06	(0.00)	0.000	0.00	(0.00)	0.553	0.05	(0.01)	0.45	(0.00)
	Library Media Specialist	−0.01	(0.03)	0.696	−0.04	(0.03)	0.173	0.01	(0.03)	0.635	0.03	(0.01)	0.34	(0.02)
	School Services Personnel	0.04	(0.04)	0.325	−0.08	(0.04)	0.048	−0.01	(0.04)	0.750	0.05	(0.01)	0.30	(0.01)
Professionalism					−0.04	(0.02)	0.010				0.08	(0.01)	0.36	(0.00)

Note: District was specified as a random effect in the regression model. The four categories (Informal/unannounced, Informal/announced, Formal/unannounced, Formal/announced) were coded into three orthogonal-contrast variables: (1) Formal vs. Informal, (2) Announced vs. Unannounced, and (3) Interaction effect. Statistically significant effects highlighted in yellow.

## Figure 2. Mean Planning Rubric Score by Announced Status and Formality for General Educators

Planning, General Educator
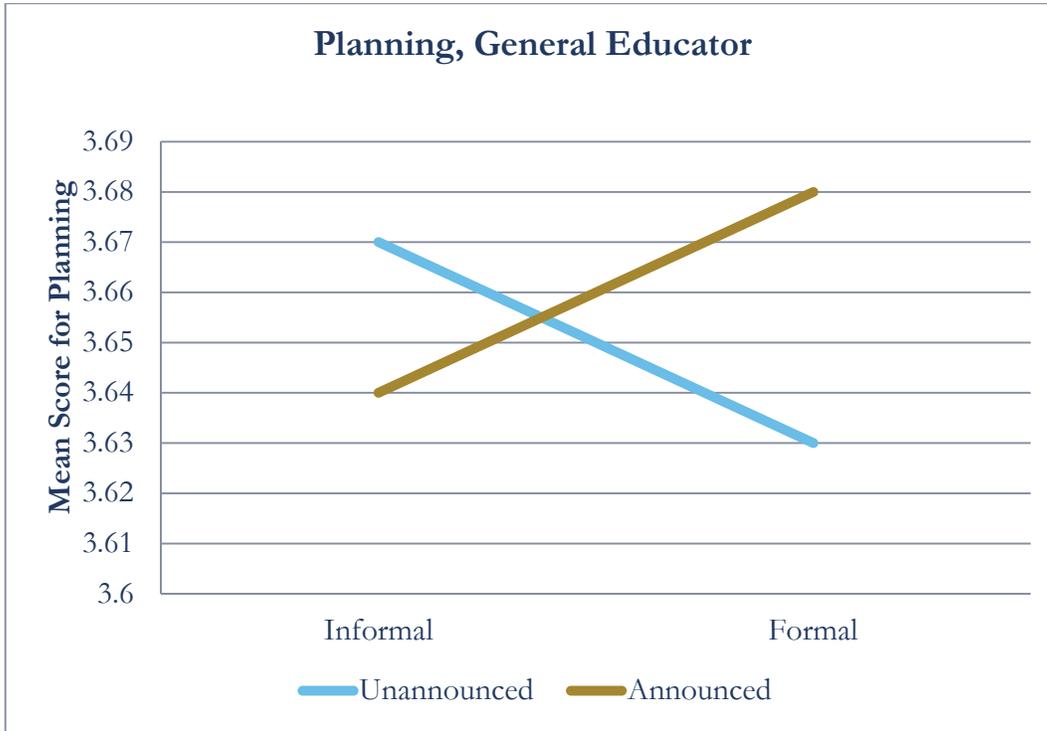
Figure 3. Mean Environment Rubric Score by Announced Status and Formality for General Educators

Environment, General Educator
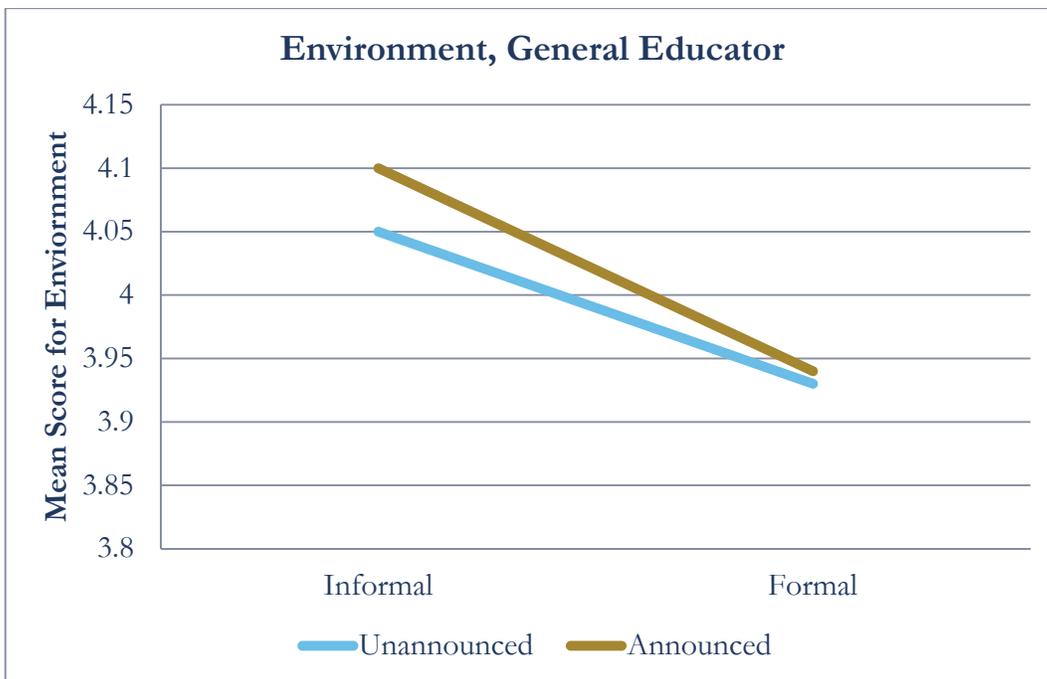
Figure 4. Mean Environment Rubric Score by Announced Status and Formality for School Services Personnel

Environment, School Services Personnel
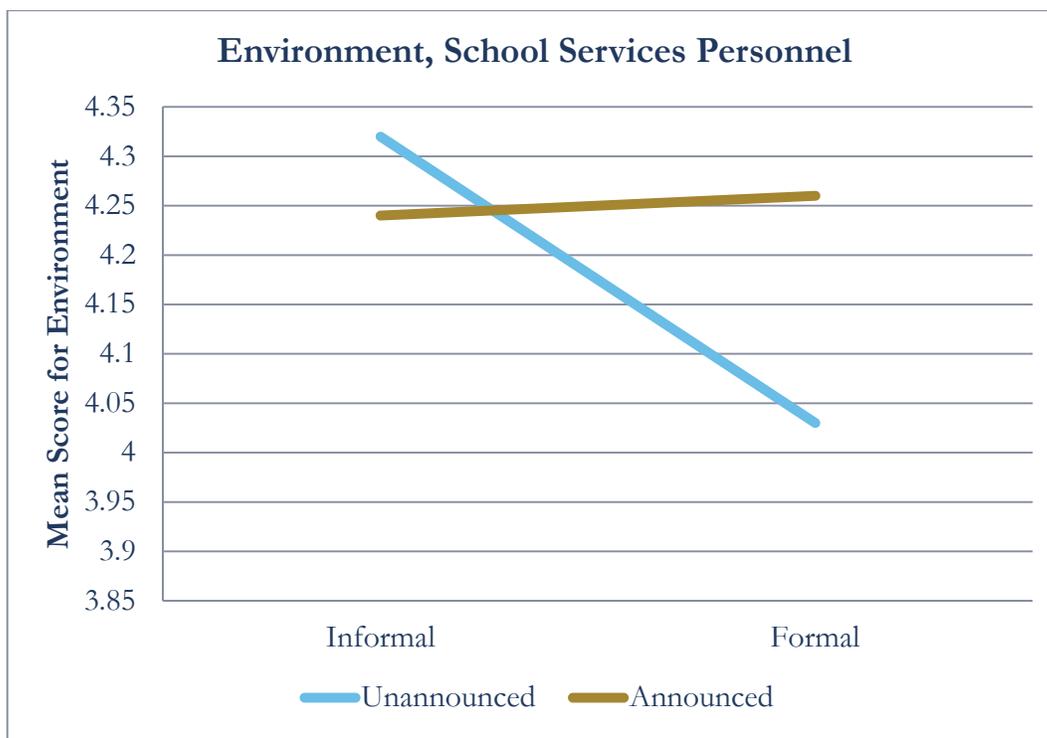
The results displayed in Figure 2, the plot for Planning by General Educators in the random-effects regression seems at odds with the simple *t*-test results discussed before. This can be understood by noting that the cells vary widely in sample size, so the natural tendency to assign equal weight to each of cells the plot leads to an erroneous impression.

For completeness, the last row of Table 4 also presents a final analysis was done using scores on Professionalism dimension. Note that for Professionalism, only Announced-Formal and Unannounced-Informal observations were present. The mean for formal-announced observations had a lower mean (3.94) than informal-unannounced observations (4.09). The effect size is 0.21. Because the two effects are confounded, it is difficult to interpret the difference.

SUMMARY FOR ANNOUNCED/UNANNOUNCED OBSERVATIONS

Overall, we found small, but significant, differences between the mean scores for Announced and Unannounced observations. Two features of this analysis should be borne in mind. Most important, although the results were for type of observation statistically significant, the actual differences in means are numerically quite small. The significance is driven more by the large number of teachers in the sample than the size of the difference in scores.

Second, examination of Table 3 shows that the number of people per condition varies substantially across conditions. This variability likely is not random; observers decided what type of observation to conduct. This may reflect focusing Formal observations on aspects of concern or some other strategy. Such a strategy would lead to different types of teachers being included in each condition, and so produce different means more as a function of who is observed, rather than due to the type of observation. The differential is borne out in the data—for the Instruction rubric, about half of the observations were Announced and the other half Unannounced. In contrast, for the Planning rubric, 6% of the observations were Announced, while 94% were Unannounced. Also recall that Professional General Educators composed the bulk (85%) of the observations for Instruction, while most (74%) of the observations for Planning were performed on Apprentice General Educators. The structure in these types of observation presumably is not random. Should future work extend these results, knowing the circumstances/decision making by which observations are announced v. unannounced and formal v. informal would be useful to contextualize and understand these results.

RUBRIC REDUCTION: PART 1, DATA ANALYSES

The goal of this task was to investigate whether mining of the 2011-2012 TEAM data would point toward reductions in the complexity of the rubric. Specifically, we investigated:

- How reliable is the summary (simple total) of the indicators in the each of the rubrics as they currently exist?
- The structure of the dimensions—do the indicators within a rubric seem to be consistently measuring the same underlying variable or is there evidence of one or more sub-dimensions?
- Which indicators are most correlated with the remainder of the indicators? Indicators with low unique variability contribute less to the overall score, and so are candidates from removal.
- What is the effect of removing indicators with the lowest unique variability? How much does this degrade the reliability of the final sum of the indicators left?

The analyses presented in this section do not account for clustering in the data. As the analyses were largely exploratory and did not include significance testing, it was deemed more useful to keep the analysis (relatively) simple[3]. If interesting effects were identified, additional analyses could be conducted that account for clustering effects.

RELIABILITY OF THE CURRENT RUBRICS

To facilitate the discussion, Table 5 shows the current allocation of individual indicators (items) to rubrics.

---

[3] This decision was also driven, in part, by limitations of readily available software. Estimation of hierarchical effects in reliability coefficients and factor analysis would entail substantial additional complication using specialized estimation software, while generating limited additional information.

Table 5. Indicator-rubric Assignment by Observation Types in the 2011-2012 TEAM Data

Rubric/Indicator	Observation Type				Abbrev
	General Educator	Library Media Specialist	School Services Personnel	Educational Professional	
Instruction/Delivery of Service					
Standards and Objectives	X	X	X		SO
Motivating Students	X	X	X		MS
Presenting Instructional Content	X	X			PIC
Information Literacy Skills		X			ILS
Lesson Structure and Pacing	X	X			LS
Activities and Materials	X	X	X		ACT
Questioning	X	X			QU
Academic Feedback	X	X			FEED
Grouping Students	X				GRP
Teacher Content Knowledge	X				TCK
Teacher Knowledge of Students	X				TKS
Thinking	X	X			TH
Problem Solving	X		X		PS
Monitoring Student Understanding		X			MON
LMS/LIS Content Knowledge		X			CK
LMS/LIS Knowledge of Students		X	X		KS
Delivery of Professional Services			X		DPS
Service Structure and Planning			X		SS
Communication			X		COM
Consultation			X		CON
Developing Educational Plans for Students			X		DEV
Organization of Services			X		OS

Planning					
Instructional Plans	X				IP
Student Work	X				SW
Assessment	X				AS
Media Center Management		X			MCM
Media Center Resources		X			MCR
Media Center Collaboration		X			MCC
Scope of Work			X		SOW
Analysis of Work Products			X		AWP
Evaluation of Services or Programs			X		EVAL
Environment					
Expectations	X	X	X		EX
Managing Student Behavior	X	X	X		MSB
Environment	X	X	X		ENV
Respectful Culture	X	X	X		RC
Professionalism					
Professional Development Opportunities				X	PROF01
Implementing New Strategies				X	PROF02
Personal Learning Plan				X	PROF03
Self-assessment				X	PROF04
Improve Performance				X	PROF05
Using Student Achievement Data				X	PROF06
Supporting School Activities and Events				X	PROF07
Adherence to Personnel Policies				X	PROF08
Safe and Orderly Learning Environment				X	PROF09
Contribute to School Community				X	PROF10

Note: Professional content knowledge (PCK; School Services Personnel observation type in the Instruction / Delivery of Service subcategory) was removed from the rubric due to insufficient observations in the data file (only 13 records).

Table 6 shows basic descriptive statistics for each rubric, disaggregated by Professional being observed. The vast majority of the observations are from observations of General Educators. For the purposes of analyses, the sample sizes for General Educators and for School Services Personnel easily meet general recommendations for the techniques (especially factor analysis) discussed below. The smallest sample size (for Library/Media Specialist for Planning and Environment) meets the more conservative recommendations for factor analysis (minimum of 1000 cases), so analyses were conducted for all of the types of observation for each of the rubrics.

Table 7 shows the proportion of observations from each rubric. There are no teachers with ratings for Instruction/Delivery of Service as well as observation scores for any of the other rubrics.[4] Similarly, the Professionalism rubric observations do not have data overlap with the other three rubrics. Finally, the overlap of the Planning and Environment rubrics (labeled "Planning + Env." in the table) is a very small proportion of the data, comprising only 1-5% of the observations for General Educators, and substantially less for the other types of educator. Given the small amount of overlap, and possible unrepresentativeness of the cases with observation data on both rubrics, it was not appropriate to perform analyses combining the indicators of Planning with the indicators of Environment.

Given very little or no data across rubrics, we were not able to investigate combinations of the rubric dimensions, nor were we able to investigate the effect of moving indicators between rubrics, based on data structures. As a result, results will be presented for each of the dimensions separately.

---

[4] All statements about data available are relative to the data set we were provided. Teachers may well have Observations for both Instruction/Delivery of Service, but we were not able to match the observations using the 20122-2012 data.

Table 6. Descriptive Statistics of the TEAM Rubric by Observation Type

Rubric	Observation	n	Mean	SD	Quartiles				
					Min	First	Median	Third	Max
Instruction / Delivery of Service	General Educator	105,509	3.65	0.62	1.00	3.25	3.67	4.08	5.00
	Library Media Specialist	2,170	3.82	0.59	1.08	3.42	3.83	4.25	5.00
	School Services Personnel	8,333	4.10	0.57	1.18	3.73	4.18	4.55	5.00
Planning	General Educator	57,871	3.64	0.71	1.00	3.00	3.67	4.00	5.00
	Library Media Specialist	1,143	4.15	0.69	1.67	3.67	4.33	4.67	5.00
	School Services Personnel	4,614	4.08	0.66	1.33	3.67	4.00	4.67	5.00
Environment	General Educator	58,247	4.03	0.70	1.00	3.50	4.00	4.50	5.00
	Library Media Specialist	1,106	4.21	0.61	2.00	3.75	4.25	4.75	5.00
	School Services Personnel	4,598	4.32	0.58	1.50	4.00	4.50	4.75	5.00
Professionalism		53,107	4.08	0.64	1.00	3.60	4.10	4.60	5.00

Table 7. Percent of Sample by Observation Type

Announced?	Subcategory Rubric	Observation Type (%)				Total
		General Educator	Library Media Specialist	School Services Personnel	Educational Professional	
Unannounced	Instruction	32.4	0.6	4.8	0.0	**37.8**
	Planning	0.8	0.0	2.4	0.0	**3.2**
(*n*=165,663;	Planning + Env.	1.2	0.0	0.4	0.0	**1.6**
58.7%)	Environment	27.2	0.5	2.3	0.0	**30.0**
	Professionalism	0.0	0.0	0.0	27.4	**27.4**
	Total	**61.5**	**1.2**	**9.8**	**27.4**	**100.0**
Announced	Instruction	40.9	0.9	0.2	0.0	**42.0**
	Planning	40.9	0.9	0.0	0.0	**41.8**
(*n*=116,480;	Planning + Env.	5.8	0.1	0.0	0.0	**5.9**
41.3%)	Environment	3.5	0.1	0.1	0.0	**3.7**
	Professionalism	0.0	0.0	0.0	6.6	**6.6**
	Total	**91.1**	**2.0**	**0.3**	**6.6**	**100.0**
Combined	Instruction	35.9	0.8	2.9	0.0	**39.5**
	Planning	17.4	0.4	1.4	0.0	**19.1**
(*n*=282,143;	Planning + Env.	3.1	0.0	0.2	0.0	**3.4**
100.0%)	Environment	17.4	0.4	1.4	0.0	**19.2**
	Professionalism	0.0	0.0	0.0	18.8	**18.8**
	Total	**73.8**	**1.5**	**5.9**	**18.8**	**100.0**

Note: For the "Planning + Env." rubric, teachers were rated on both the Planning and Environment rubric.

DATA STRUCTURE OF INSTRUCTION/DELIVERY OF SERVICE ("INSTRUCTION")

Analysis

Overall, the reliability (as measured by Cronbach's alpha) was found to be ~0.92 for all three of the observation types using the Instruction rubric. This indicates a very high level of internal consistency. Table 8 presents correlation analysis for the Instruction rubric. Analyses were performed separately for each type of observation, as indicated by the heavier line separating the sections of the table. For the instruction rubric, indicator-rest correlations were above 0.6 and average inter-indicator correlations were above 0.5.

The second step in the analysis of the Instruction rubric was to perform exploratory factor analysis (EFA). The analysis is based on Pearson correlations.[5] The goal of this analysis was to see if any subgroups of indicators could be detected. Initial eigenvalues were extracted, and a scree plots was created for each observation type (see Figures 5 through 7). Based on the scree plot, one or possibly two factors are consistent with the data. The exploratory factor analysis was performed using maximum likelihood factor analysis, and the subjecting the loadings to an oblique rotation (promax).[6] Table 9 presents the factor pattern matrix for the 2-factor solution for each observation type.

For General Educators and School Services Personnel, the second factor has only two substantial (>0.5) loadings. Such a factor (referred to as a "doublet") is generally the result of a higher than expected correlation between a single pair of variables (Thinking & Problem Solving for General Educators, and Knowledge of Students and Motivating Students for School Services Personnel). Doublet factors are indications extracting too many factors, leading us to conclude that a single factor solution is more appropriate. For Library/Media Specialist, the factor pattern is somewhat different; all but 3 indicators have loadings > 0.5 on the second factor. The correlation between factor 1 and factor 2 is 0.74, indicating that the two factors share more than half their variance. Again, we decided the 1-factor solution was the most representative of the data.

---

[5] Pearson correlations are the strength of the association between the observed scores 1-5. Factor analysis of the polychoric correlations was also performed. A polychoric correlation assumes that two normally distributed variables underlie the observed scores. The observed score is then obtained by setting a series of cut points, then assigning the integer score based on which cut points the latent variable exceeds. The polychoric correlation is based on a model of underlying relationship. As our focus here is to make inferences about the total score of observed ratings, we chose the Pearson correlation as more appropriate. For all four rubrics and all three types of observation, analysis of the polychoric correlations indicated a 1-factor solution.

[6] We extracted and present results for two factors for Instruction and for the Professionalism rubric. The scree plot is criterion is somewhat subjective, so we chose to extract more factors than might strictly be needed. We felt that examining the putative second factor was useful, although we eventually chose a 1-factor solution. Additional analyses of polychoric correlations (not shown) supported this decision (see footnote 4).

Table 8. Instruction Rubric: Correlation Analysis

Observation	Indicator	Correlation			Cronbach's alpha
		Indicator-test	Indicator-rest	Inter-indicator	
General Educator	SO	0.73	0.67	0.51	0.92
	MS	0.76	0.71	0.51	0.92
	PIC	0.79	0.74	0.50	0.92
	LS	0.75	0.69	0.51	0.92
	ACT	0.76	0.71	0.51	0.92
	QU	0.73	0.67	0.51	0.92
	FEED	0.73	0.68	0.51	0.92
	GRP	0.71	0.64	0.52	0.92
	TCK	0.75	0.69	0.51	0.92
	TKS	0.73	0.67	0.51	0.92
	TH	0.75	0.69	0.51	0.92
	PS	0.72	0.66	0.51	0.92
Library Media Specialist	SO	0.72	0.65	0.50	0.92
	MS	0.73	0.67	0.50	0.92
	PIC	0.77	0.72	0.50	0.92
	ILS	0.70	0.64	0.51	0.92
	LS	0.73	0.67	0.50	0.92
	ACT	0.75	0.70	0.50	0.92
	QU	0.73	0.66	0.50	0.92
	FEED	0.74	0.68	0.50	0.92
	TH	0.73	0.67	0.50	0.92
	MON	0.79	0.74	0.49	0.91
	CK	0.73	0.67	0.50	0.92
	KS	0.73	0.67	0.50	0.92
School Services Personnel	SO	0.74	0.68	0.50	0.91
	MS	0.72	0.65	0.50	0.91
	ACT	0.71	0.64	0.50	0.91
	PS	0.72	0.66	0.50	0.91
	KS	0.70	0.63	0.50	0.91
	DPS	0.77	0.71	0.49	0.91
	SS	0.76	0.70	0.49	0.91
	COM	0.75	0.69	0.49	0.91
	CON	0.75	0.68	0.49	0.91
	DEV	0.71	0.64	0.50	0.91
	OS	0.77	0.71	0.49	0.91

Note: "Indicator-test" correlation indicates association between the individual indicator and the overall measure. "Indicator-rest" correlation (or adjusted correlation) is the association between the indicator and the measure excluding the indicator being correlated. "Inter-indicator" correlation shows the average inter-indicator correlations. "Cronbach's alpha" represents internal-consistency reliability when the individual indicator is removed. Indicators considered for removal highlighted.
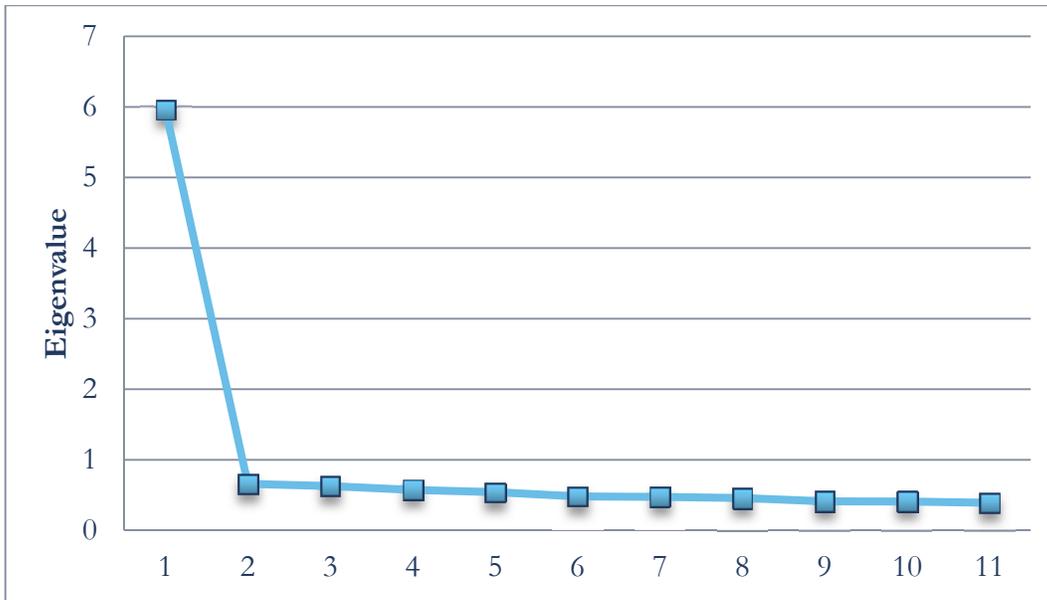
Figure 5. Scree Plot for Instruction Rubric, General Educator
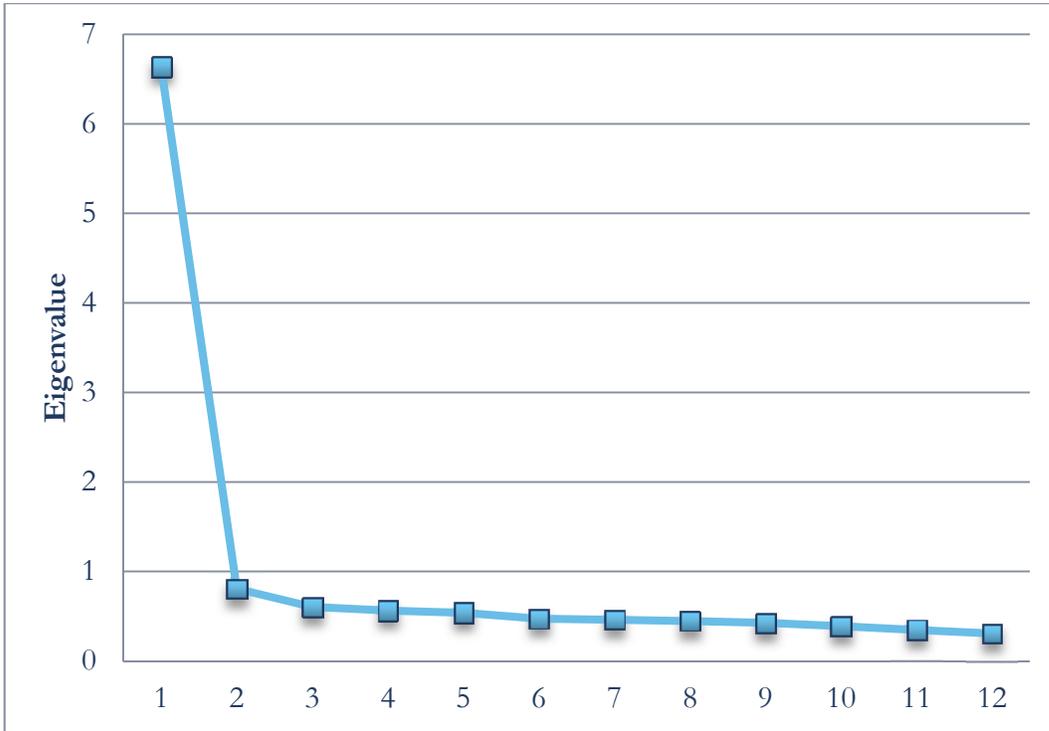
Figure 6. Scree Plot for Instruction Rubric, Library Media Specialist
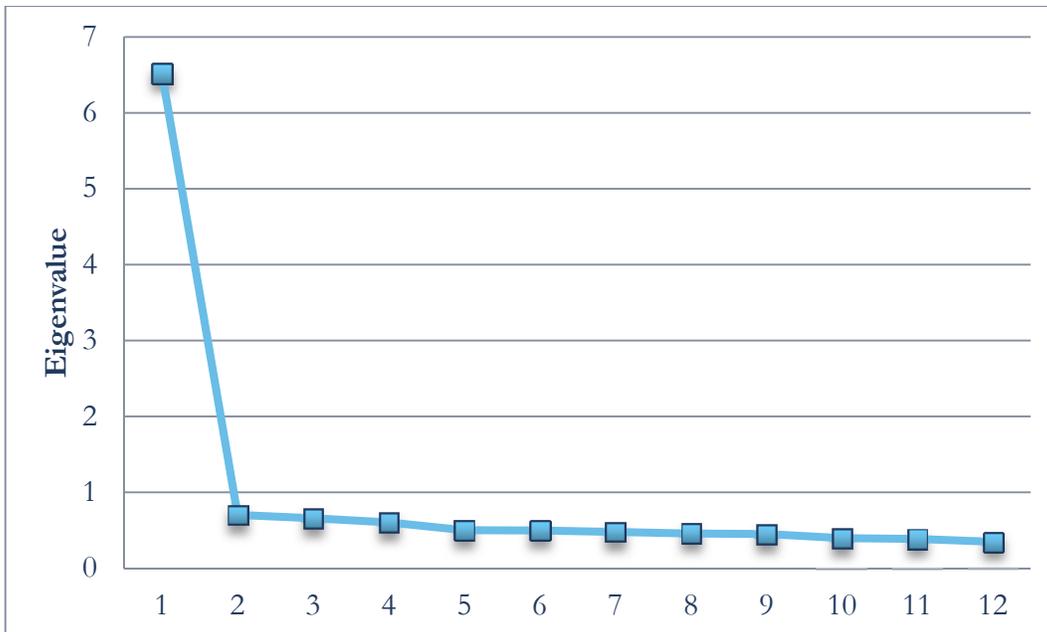
**Figure 7. Scree Plot for Instruction Rubric,
School Service Personnel**

Table 9: Exploratory Factor Analysis (EFA) for Instruction Rubric

Observation	Indicator	2-Factor Solution			1-Factor Solution	
		Factor 1	Factor 2	Unique Variance	Factor 1	Unique Variance
General Educator	SO	0.65	0.07	0.50	0.70	0.51
	MS	0.65	0.12	0.45	0.74	0.45
	PIC	0.73	0.07	0.38	0.78	0.40
	LS	0.74	0.01	0.44	0.73	0.47
	ACT	0.56	0.22	0.46	0.74	0.46
	QU	0.39	0.37	0.51	0.70	0.52
	FEED	0.56	0.18	0.51	0.70	0.51
	GRP	0.62	0.08	0.54	0.67	0.55
	TCK	0.64	0.13	0.47	0.73	0.47
	TKS	0.70	0.02	0.48	0.70	0.51
	TH	0.02	0.85	0.26	0.72	0.48
	PS	0.06	0.76	0.36	0.69	0.52
Library Media Specialist	SO	0.63	0.07	0.52	0.68	0.54
	MS	0.73	-0.01	0.48	0.70	0.51
	PIC	0.75	0.03	0.41	0.75	0.44
	ILS	0.27	0.47	0.52	0.67	0.55
	LS	0.63	0.11	0.50	0.70	0.50
	ACT	0.64	0.12	0.47	0.73	0.47
	QU	0.36	0.40	0.50	0.69	0.52
	FEED	0.55	0.20	0.48	0.72	0.48
	TH	-0.03	0.85	0.31	0.69	0.52
	MON	0.59	0.22	0.41	0.77	0.41
	CK	0.55	0.17	0.52	0.69	0.52
	KS	0.49	0.25	0.52	0.70	0.52
School Services Personnel	SO	0.54	0.20	0.50	0.71	0.50
	MS	0.00	0.77	0.41	0.67	0.55
	ACT	0.43	0.28	0.55	0.67	0.55
	PS	0.50	0.22	0.53	0.69	0.53
	KS	0.18	0.53	0.53	0.65	0.58
	DPS	0.61	0.17	0.44	0.75	0.44
	SS	0.59	0.18	0.46	0.74	0.46
	COM	0.56	0.20	0.47	0.73	0.47
	CON	0.56	0.20	0.48	0.72	0.48
	DEV	0.47	0.24	0.55	0.67	0.55
	OS	0.83	-0.06	0.39	0.75	0.44

Note: Results from "EFA" indicate exploratory factor analysis. Factor loadings based on oblique rotation (promax). Indicators considered for removal highlighted.

Instrument Reduction

Table 9 also displays the estimated uniqueness for each indicator. In factor analysis, uniqueness is the variance in the indicator score that is not accounted for by the factors. Based upon the factor analysis results, candidate indicators were selected for deletion. These indicators had the lower uniqueness values. The last column of the Table 8 shows the value of Cronbach's alpha if the indicator were deleted from the rubric. For most indicators, the effect is less than a 0.01 decrease in reliability. In no case is the decrease as large as 0.02. This process was iterated until a set of indicators (shown by yellow highlighting) for deletion was identified. Based on correlations, factor loadings, unique variance contributions, and changes in Cronbach's alpha, indicators with high associations and low information were flagged.

For the General Educator observations, 3 indicators (PIC, LS, TKS) were selected for deletion. Likewise, the 3 candidate indicators were selected for the Library Media Specialist observations (indicators MS, PIC, ACT) and 3 indicators were selected for the School Service Personnel (MS, DPS, OS). The indicators in each reduced rubric were used to create mean ratings and were compared to the original mean ratings. Correlations between the reduced and original mean ratings were all above 0.95. There were minor changes in Cronbach's alpha between the original and reduced measures for Instruction (less than 0.05 point difference). The final recommendations and the changes in coefficient alpha are summarized in the top section of Table 13.

PLANNING RUBRIC

Analysis

The same set of analyses was performed for the Planning rubric. Table 10 shows (in a more compressed format) the results for the analysis of the planning rubric. Because the Planning rubric for General Educator and School Service Personnel had only 3 indicators, factor analysis was unlikely to identify meaningful subgroups of indicators.[7] For Library Media Specialist observations, the four indicators also indicated a 1-factor solution. Because of the limited number of indicators, scree plots are not presented for factor analysis of the Planning rubric.

Instrument Reduction

The chief information of interest from the EFA was the uniqueness of each variable. However, the final decision of which indicator(s) to exclude from each of the observation types was largely based on the change in coefficient alpha. One indicator (SW) was selected for the General Educator observation type. Similarly, one indicator (MCM) was selected for the Library Media Specialist observation type, and one (AWP) for the School Service Personnel observation type.

The indicators in each reduced rubric were used to create mean ratings and were compared to the original mean ratings. Correlations between the reduced and original mean ratings were all above 0.95. There were minor changes in Cronbach's alpha between the original and reduced measures for

---

[7] Indeed, most factor analysis texts recommend 3 indicator variables (indicators in the present context) be used to identify *each factor*.

Planning was larger than that for the Instruction rubric, but the difference in alpha was less than 0.12 in each case. The final recommendations and the changes in coefficient alpha are summarized in the second section of Table 13.

ENVIRONMENT

Analysis

Because the Environment rubric had only 3 indicators, factor analysis was unlikely to identify meaningful subgroups of indicators. The indicators did indicate a single factor. Table 11 shows the results for the correlation analysis and EFA of a 1-factor solution for the Environment rubric. The criterion of interest from the factor analysis was the uniqueness of each variable. Due to the limited number of indicators for the Environment rubric, creating scree plots for the EFA were not necessary. But as was the case for the Planning rubric, the final decision of which indicator(s) to exclude from each of the observation types was largely based on the change in Cronbach's alpha.

Instrument Reduction

For each observation type, indicator RC was selected as the best candidate for removal. There were moderate changes in Cronbach's alpha between the original and reduced measures for the Environment rubric was larger than that for the Instruction rubric, but the difference was less than 0.12 in each case. The final recommendations and the changes in coefficient alpha are summarized in the third section of Table 13.

PROFESSIONALISM

Analysis

Recall that the Professionalism rubric is not different based upon the type of observation being performed; in our data it corresponded to a separate, 4th kind of observation. There are 10 indicators on the Professionalism rubric. Overall, the internal consistency measure of reliability (Cronbach's alpha) is very high (0.92) for Professionalism. Table 12 presents correlation analysis for the Professionalism rubric. Indicator-rest correlations for the Professionalism rubric were above 0.6 and average inter-indicator correlations were above 0.5.

The second step in the analysis of the Professionalism rubric was to perform exploratory factor analysis (EFA). The analysis is based on Pearson correlations. The goal of this analysis was to see if any subgroups of indicators could be detected. Initial eigenvalues were extracted, and a scree plot was created (see Figure 9). Based on the plot, 1 or 2 factors are consistent with the data.

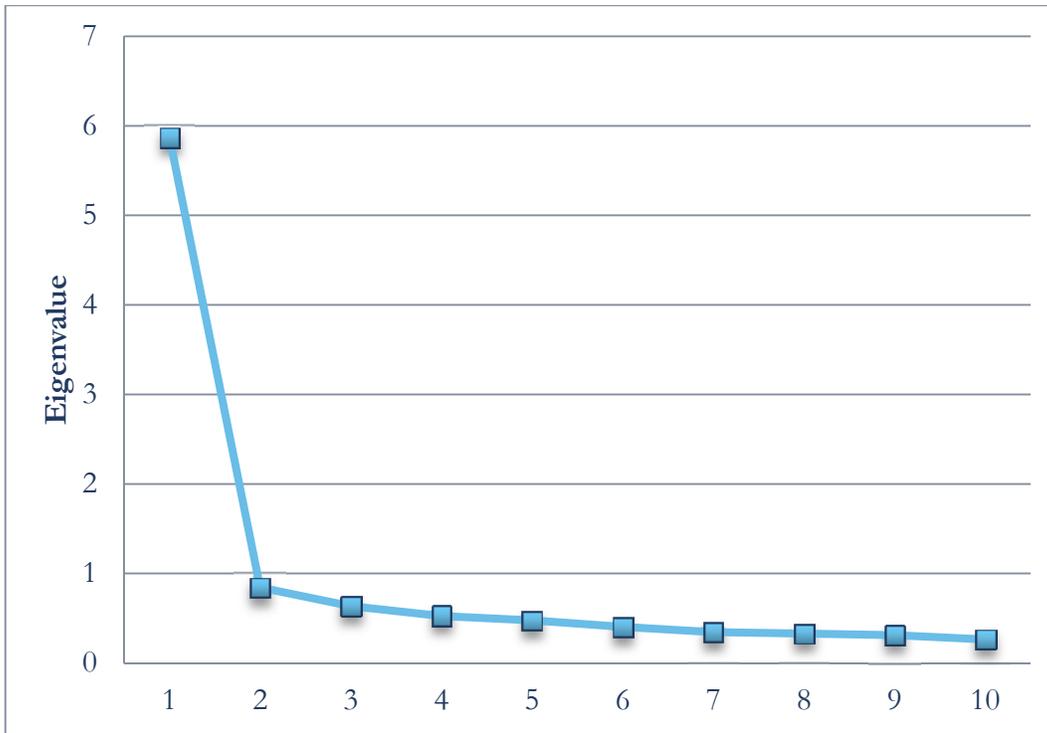**Figure 9. Scree Plot for Analysis of Professionalism Rubric**

Table 12 also presents the factor pattern matrix, using maximum likelihood factor analysis, and then subjecting the loadings to an oblique rotation (promax) for the 2-factor solution. The second factor has only two substantial (>0.5) loadings (PROF8 and PROF9), and so could be considered a doublet. However, PROF07 has a loading > 0.4, so the case is less clear. As noted above, doublet factors are indications extracting too many factors. The estimated correlation between the factors is also fairly high (0.74), leading us to conclude that a single-factor solution is more appropriate.

Instrument Reduction

Table 12 also displays the estimated uniqueness for each indicator. In factor analysis, uniqueness is the variance in the indicator score that is not accounted for by the factors. Based upon the factor analysis results, candidate indicators were selected for deletion. These indicators had the lower uniqueness values. The last column of Table 12 shows the value of Cronbach's alpha if the indicator were deleted from the rubric. For most indicators, the effect is less than a 0.01 decrease in reliability. In no case is it as large as 0.02. This process was iterated until a set of indicators (shown by yellow highlighting) for deletion was identified. Based on correlations, factor loadings, unique variance contributions, and changes in Cronbach's alpha, indicators with high associations and low information were flagged.

Three indicators (PROF02, PROF03, and PROF05) were selected for deletion. The indicators in each reduced rubric were used to create mean ratings and were compared to the original mean ratings. Correlations between the reduced and original mean ratings were all above 0.95. There were

minor changes in Cronbach's alpha between the original and reduced measures for the Professionalism rubric (less than 0.05 point difference). The final recommendations are summarized in the last row of Table 13.

Table 10. Instrument Reduction Data Analysis for Planning Rubric: Correlation and Exploratory Factor Analysis

Observation	Indicator	Correlation			EFA		Cronbach's alpha
		indicator-test	indicator-rest	inter-indicator	Factor loading	Unique Variance	
General Educator	IP	0.87	0.70	0.64	0.79	0.37	0.78
	SW	0.87	0.70	0.64	0.80	0.37	0.78
	AS	0.87	0.71	0.63	0.81	0.35	0.77
Library Media Specialist	MCM	0.88	0.72	0.60	0.83	0.31	0.75
	MCR	0.87	0.71	0.62	0.81	0.34	0.76
	MCC	0.85	0.67	0.67	0.74	0.45	0.81
School Services Personnel	SOW	0.83	0.60	0.55	0.72	0.49	0.71
	AWP	0.85	0.65	0.49	0.80	0.36	0.66
	EVAL	0.82	0.59	0.57	0.69	0.53	0.73

Note: "Indicator-test" correlation indicates association between the individual indicator and the overall measure. "Indicator-rest" correlation (or adjusted correlation) is the association between the indicator and the measure excluding the indicator being correlated. "Inter-indicator" correlation shows the average inter-indicator correlations. "Cronbach's alpha" represents internal-consistency reliability when the individual indicator is removed. Indicators considered for removal highlighted.

Table 11. Instrument Reduction Data Analysis for Environment Rubric: Correlation and Exploratory Factor Analysis

Observation	Indicator	Correlation			EFA		Cronbach's apha
		indicator-test	indicator-rest	inter-indicator	Factor loading	Unique Variance	
General Educator	EX	0.82	0.67	0.59	0.75	0.44	0.81
	MSB	0.85	0.70	0.56	0.80	0.36	0.79
	ENV	0.80	0.63	0.61	0.70	0.51	0.82
	RC	0.85	0.71	0.55	0.81	0.35	0.79
Library Media Specialist	EX	0.80	0.63	0.50	0.73	0.46	0.75
	MSB	0.80	0.63	0.51	0.73	0.47	0.75
	ENV	0.76	0.57	0.54	0.65	0.58	0.78
	RC	0.82	0.66	0.48	0.76	0.43	0.74
School Services Personnel	EX	0.80	0.63	0.50	0.73	0.47	0.75
	MSB	0.79	0.60	0.52	0.69	0.52	0.76
	ENV	0.79	0.62	0.51	0.71	0.49	0.76
	RC	0.80	0.62	0.50	0.72	0.48	0.75

Note: "Indicator-test" correlation indicates association between the individual indicator and the overall measure. "Indicator-rest" correlation (or adjusted correlation) is the association between the indicator and the measure excluding the indicator being correlated. "Inter-indicator" correlation shows the average inter-indicator correlations. "Cronbach's alpha" represents internal-consistency reliability when the individual indicator is removed. Results from "EFA" indicate exploratory factor analysis. Factor loadings based on oblique rotation (promax). Indicators considered for removal highlighted.

Table 12. Instrument Reduction Data Analysis for Professionalism: Correlation and Exploratory Factor Analysis

Indicator	Correlation			2-Factor Solution			1-Factor Solution		Cronbach's alpha
	Indicator-test	Indicator-rest	Inter-indicator	Factor 1	Factor 2	Unique Variance	Factor 1	Unique Variance	
PROF01	0.75	0.68	0.54	0.41	0.35	0.50	0.70	0.51	0.91
PROF02	0.82	0.77	0.52	0.78	0.06	0.31	0.82	0.33	0.91
PROF03	0.81	0.76	0.53	0.81	0.02	0.32	0.81	0.35	0.91
PROF04	0.82	0.77	0.52	0.74	0.11	0.32	0.82	0.33	0.91
PROF05	0.83	0.78	0.52	0.75	0.12	0.30	0.83	0.30	0.91
PROF06	0.76	0.69	0.54	0.73	0.03	0.43	0.74	0.45	0.91
PROF07	0.61	0.52	0.57	0.12	0.47	0.69	0.52	0.73	0.92
PROF08	0.75	0.68	0.54	0.05	0.76	0.37	0.69	0.53	0.91
PROF09	0.76	0.69	0.54	0.05	0.77	0.34	0.70	0.51	0.91
PROF10	0.73	0.66	0.54	0.38	0.34	0.54	0.67	0.55	0.91

Note: "Indicator-test" correlation indicates association between the individual indicator and the overall measure. "Indicator-rest" correlation (or adjusted correlation) is the association between the indicator and the measure excluding the indicator being correlated. "Inter-indicator" correlation shows the average inter-indicator correlations. "Cronbach's alpha" represents internal-consistency reliability when the individual indicator is removed. Results from "EFA" indicate exploratory factor analysis. Factor loadings based on oblique rotation (promax). Indicators considered for removal highlighted.

Table 13. Indicators in the Reduced Rubrics

Rubric	Observation type	Number of Indicators		Indicators Dropped in Reduced Rubric
		Original	Reduced	
Instruction	General Educator	12	9	PIC, LS, TKS
	Library Media	12	9	MS, PIC, ACT
	School Services	11	8	MS, DPS, OS
Planning	General Educator	3	2	SW
	Library Media	3	2	MCM
	School Services	3	2	AWP
Environment	General Educator	4	3	RC
	Library Media	4	3	RC
	School Services	4	3	RC
Professionalism		10	7	PROF02, PROF03, PROF05

Note: Indicators indicated in the column "Indicators Dropped in Reduced Rubric" are the indicators that were highlighted in Tables 9 through 13. These indicators have high indicator-rest correlation, inter-indicator correlation, and high factor loadings; they also have relatively low unique variance.

SUMMARY OF INDICATOR DELETION RECOMMENDATIONS

Overall, statistical analysis identified three indicators (for each type of Observation) for the Instruction rubric, one for each type of Observation for Planning and one indicator for each type of Observation for Environment. We also identified three indicators for the Professionalism rubric. Table 14 summarizes some statistical evidence about the relationship between the original and reduced forms of each rubric. The column labeled "Correlation with Indicators Removed" is the correlation between two total scores for the rubric. The first score is uses all indicators in the original rubric. The second score uses the reduced rubric. In every case the correlation is greater than or equal to 0.95. The last two columns of the table show coefficient alpha for the original and reduced forms of the rubric. In most cases the decrease is minor; the exception is the Planning rubric for School Services Personnel, where the decline is 0.12.

Given the limited number of indicators (3-4) reduction of the Planning and Environment rubrics, the advantage of decreasing the number of indicators does not seem warranted. Our analyses indicate that the Instruction/Delivery of Service rubric and the Professionalism rubric can be shorted by 25-30% without substantial decrease in the reliability of scores on the overall rubric.

Examination of the value of the alpha coefficient with the indicator removed in Table 8 for the Instruction/Delivery of Service rubric, and Table 12 for the Professionalism rubric, shows little difference among the indicators. Thus, the choice of which specific indicators to remove was somewhat arbitrary, and there is the real possibility that the selection may capitalize on chance variation within this data set. Given these considerations, we conclude that there is not sufficient

evidence to support a data-driven modification of the rubrics. Our final recommendation is that theory-based modifications (presented in the next section) be given preference.

Table 14. Changes in Reliability Cronbach's alpha with Rubric Reduction

Rubric	Observation Type	Correlation Full Score with Reduced Score (Indicators Removed)	Cronbach's alpha	
			Original	Reduced
Instruction	General Educator	0.97	0.92	0.88
	Library Media	0.97	0.92	0.87
	School Services	0.97	0.92	0.86
Planning	General Educator	0.96	0.84	0.77
	Library Media	0.97	0.84	0.75
	School Services	0.95	0.78	0.66
Environment	General Educator	0.98	0.85	0.79
	Library Media	0.97	0.81	0.74
	School Services	0.97	0.80	0.75
Professionalism		0.97	0.92	0.87

RUBRIC REDUCTION: PART 2, SUBSTANTIVE REVIEW AND RECOMMENDATIONS

The third component of this review of the TEAM instrument was a substantive review. Two experts in teaching practice evaluation were asked to review the current version of the TEAM instrument. The experts independently reviewed the instrument, and recorded suggestions for improvement. Then the two exchanged lists, and met to reconcile the differences between their versions. The result is a final list of suggestions for modification to each of the rubrics.

The experts' final recommendations are presented in two sections—general, overarching observations, and specific recommendations at the rubric/indicator level.

The general remarks are summarized here:
- Certain indicators are important measures of teacher practice. However, they were not clearly observable in Instruction. These indicators were moved to Planning where they could be more reliably observed during discussion with the teacher. Indicators of practice that belong in Planning, not in Instruction:
    - Knowledge of Content
    - Knowledge of Students
    - Grouping Students
    - Use of Data (with respect to Student Achievement data)
- Certain qualifiers (vague quantifiers like "mostly", subjective terms like "appropriate") make it difficult to reliably score teacher practice.
- Redundancy of indictors makes the tool cumbersome.
- Need to separate each indicator to insure reliability of scoring.
- Inconsistency of language across the levels of effectiveness makes it difficult to reliably rate teacher practice.
- The 5-point structure of the instrument may be problematic. Having no language describing the "4" and "2" score points makes it harder to reliably assign scores. It also makes it difficult to provide teachers with productive feedback to improve their practice. If the 5-level scale is maintained, the language for levels 4 and 2 should be added.
- A 5-level rubric may result in very many 3s and very few 1s and 5s being scored, which again may be problematic for delivering helpful feedback to teachers. An option to consider may be a 4-level scale for the instrument (e.g., "Undeveloped", "Developing", "Proficient", "Distinguished") with clear, descriptive language at each score point.
- For some of the rubrics, we suggest sub-dividing the indicators to facilitate scoring.

Table 15 summarizes the salient recommendations for the Instruction rubric. In the Appendix, Table A6 gives the line-by-line suggested revisions to the language.

Table 15. Modifications to Instruction Rubric

Instruction	
Standards and Objectives	• Deleted indicators that are difficult to observe, or are present in another component. • Language added to remaining indicators to support a more reliable assessment of practice.
Motivating Students	• Deleted this component, redundant.
Presenting Instructional Content:	• Language added to indicators to support a more reliable assessment of practice.
Lesson Structure and Pacing	• Language added to indicators to support a more reliable assessment of practice. • Deleted language that required the observed lesson to have a beginning, middle, and end; this may restrict the observation of "teachable moments".
Activities and Materials	• Use of language, "attention" deleted, how is attention in students reliably measured, is it student attention to activities and materials, or compliance? • Language added to indicators to support a more reliable assessment of practice.
Questioning	• Complete reconstruction of this component to just four indicators to support a more reliable assessment of practice.
Academic Feedback	• Deleted indicators that are difficult to observe, or are present in another component.
Grouping Students	• Deleted this component, redundant.
Teacher Content Knowledge	• Moved to planning and compressed into one component, Pedagogical Content Knowledge, that represents both Teacher Content Knowledge and Teacher Knowledge of Students. • Reconstruction of this new component to just three indicators supports a more reliable assessment of practice.
Teacher Knowledge of Students	• See above.
Thinking	• The Thinking and Problem Solving indicators are redundant. They were combined and streamlined for consistency across the levels of effectiveness.
Problem Solving	• The Thinking and Problem Solving indicators are redundant. They were combined and streamlined for consistency across the levels of effectiveness.

Table 16 summarizes the salient recommendations for the Planning rubric. In the Appendix, Table A7 gives the line-by-line suggested revisions to the language of individual indicators.

Table 16. Suggested Modifications to Indicators in Planning Rubric.

Planning Rubric	
Instructional Plans	• Replace ambiguous language with specific descriptors.
Student Work	• Reorder and replace descriptors: "create, evaluate, analyze, apply, understand information" rather than "remember". • Use above across the levels of effectiveness, with qualifiers of consistently, most, rarely.
Assessment	• Build consistency across the levels of effectiveness.

Table 17 summarizes the salient recommendations for the Environment rubric. In the Appendix, Table A8 gives the line-by-line suggested revisions to the language of individual indicators.

Table 17. Suggested Modifications to Indicators in Environment Rubric

Environment	
Expectations	• Deleted indicators that are difficult to observe, or are present in another component.
Managing Student Behavior	• Deleted indicators such as "establishing" and "communicating" the rules; may be difficult to observe. • Added language "or there is no misbehavior" to acknowledge that a lack of evidence may indicate a practice that is "Significantly Above Expectations". While this language was questioned, in the end it was viewed as a beneficial addition to improve a more reliably assessment of practice. • Questioned "attending to individual students versus class", is this language necessary. • Questioned the addition of "timely".
Environment	• Language deleted and/or modified to support a more accurate assessment of practice.
Respectful Culture	• Language deleted and/or modified to support a more accurate assessment of practice.

Table 18 summarizes the salient recommendations for the Professionalism rubric. Note that, unlike the other three rubrics, the format of Table 18 does not mirror the Professionalism rubric. The experts felt that there is little consistency across the levels of effectiveness in the current rubric. Therefore, the Professionalism rubric has been rewritten across most the indicators. The content experts also questioned how often and when this domain is scored.

Table 18. Suggested Modifications to Indicators in Professionalism Rubric

Professional	
Professional Growth and Learning	• The indicators in this component were redundant and inconsistent across the levels; therefore they were reformatted and streamlined.
Use of Data	• Language reflecting this component should be added to the Planning Component to include "Student achievement data".
School and Community Involvement	• Language was changed to build consistency across the levels of effectiveness.
Leadership	• Language was changed to build consistency across the levels of effectiveness.

In the Appendix, Table A9 gives the line-by-line suggested revisions to the language of individual indicators.

The new rubric resulting from consolidating the recommended revisions to the TEAM rubric results in a very long instrument, and one we do not suggest for use in the field. Each revised indicator is now clear, scorable, and reflects a single aspect of teaching. However, there are far too many indicators to support reliable scoring—this instrument would overwhelm even the best-trained observer. Instead, we recommend that TDOE consider a compact set of aspects of teaching practice are the most valued in Tennessee: the ones most likely to increase student achievement if teaching practice on those aspects improved. With those identified, the indicators in the revised set that assess the chosen aspects could be identified, resulting in a tightly-focused teaching practice evaluation instrument.

SUMMARY AND CONCLUSION

This paper examined investigation of the TEAM instrument, including analyses of the 2011-2012 data.

The first activity examined the effect of announced versus unannounced observations, and secondarily, formal versus informal observations. Although several comparisons reached the criterion for statistical significance, the size of the effects was quite small. Based on this data, we conclude that there is little meaningful difference between the announced and unannounced observations.

A secondary observation was that the data show evidence of clustering at the level of district. We recommend that analyses should reflect this hierarchical

The second activity was a data mining examination of the 2011-2012 observation scores. For each rubric, internal consistency reliability (Cronbach's alpha) was examined, and exploratory factor analyses were conducted to investigate the underlying dimensionality of the indicator data. The

reliability (considering number of indicators) of each of the rubrics was high. Possible reductions in each rubric were identified. However, difference in statistics for the best and worst candidate indicators for removal was small, and our recommendation is that preference be given to theory-based revisions.

The third activity was a theory-based review of the current TEAM instruments. Two content experts reviewed the instrument. Their recommendations were in two sections, general, overarching considerations and detailed suggestions at the rubric-indicator level.

We recommend that the TN DOE seriously consider the theory-based recommendations. Investing time in selection of the most highly valued aspects of teaching practice, combined with choosing aligned indicators to create a compact instrument is likely to result in a more effective evaluation system and more informative data. Should the department decide to pursue this avenue, we suggest they consider leveraging the investment of time and effort spent by the content experts who completed this analysis to understand and deeply consider existing TEAM instrument.

Appendix

Table A1. District-level Variability for the TEAM Rubric: Mixed-effects Regression

Rubric	Observation	Fixed Effect			Random Effect				Intraclass Correlation
		District (Intercept)			Var(District)		Var(Residual)		
		Est.	SE	*p*-value	Est.	SE	Est.	SE	
Instruction / Delivery of Service	General Educator	3.66	(0.02)	0.000	0.03	(0.00)	0.36	(0.00)	0.09
	Library Media Specialist	3.83	(0.03)	0.000	0.04	(0.01)	0.31	(0.01)	0.12
	School Services Personnel	4.10	(0.02)	0.000	0.04	(0.01)	0.29	(0.00)	0.13
Planning	General Educator	3.66	(0.02)	0.000	0.07	(0.01)	0.46	(0.00)	0.13
	Library Media Specialist	4.16	(0.03)	0.000	0.03	(0.01)	0.45	(0.02)	0.06
	School Services Personnel	4.08	(0.03)	0.000	0.06	(0.01)	0.39	(0.01)	0.13
Environment	General Educator	4.04	(0.02)	0.000	0.05	(0.01)	0.45	(0.00)	0.10
	Library Media Specialist	4.20	(0.03)	0.000	0.03	(0.01)	0.34	(0.02)	0.07
	School Services Personnel	4.31	(0.02)	0.000	0.05	(0.01)	0.30	(0.01)	0.14
Professionalism		4.07	(0.03)	0.000	0.08	(0.01)	0.36	(0.00)	0.18

Note: Mixed-effects regression (intercept only) with district specified as the random effect. Highlighted values represent domains/observation type with high intraclass correlation (≥ 0.10), indicating differences by districts.

Table A2. District-level Variability for Instruction/Delivery of Service: Mixed-effects Regression

Observation	Indicator	n	Fixed Effect		Random Effect				Intraclass Correlation
			District (Intercept)		Var(District)		Var(Residual)		
General Educator **(ICC: 0.09)**	SO	10,821	3.73	(0.02)	0.04	(0.01)	0.74	(0.00)	0.05
	MS	10,821	3.86	(0.02)	0.04	(0.00)	0.67	(0.00)	0.05
	PIC	10,821	3.77	(0.02)	0.04	(0.01)	0.72	(0.00)	0.05
	LS	10,821	3.62	(0.02)	0.04	(0.01)	0.73	(0.00)	0.05
	ACT	10,821	3.68	(0.02)	0.04	(0.01)	0.66	(0.00)	0.05
	QU	10,818	3.52	(0.02)	0.03	(0.00)	0.60	(0.00)	0.05
	FEED	10,815	3.57	(0.02)	0.03	(0.00)	0.60	(0.00)	0.05
	GRP	10,816	3.44	(0.02)	0.04	(0.01)	0.64	(0.00)	0.05
	TCK	10,820	4.13	(0.02)	0.04	(0.01)	0.63	(0.00)	0.07
	TKS	10,804	3.75	(0.02)	0.05	(0.01)	0.69	(0.00)	0.06
	TH	10,820	3.43	(0.02)	0.04	(0.01)	0.61	(0.00)	0.07
	PS	10,821	3.43	(0.02)	0.07	(0.01)	0.73	(0.00)	0.08
Library Media Specialist **(ICC: 0.12)**	SO	259	3.95	(0.03)	0.06	(0.02)	0.67	(0.02)	0.08
	MS	259	4.02	(0.03)	0.05	(0.01)	0.60	(0.02)	0.07
	PIC	259	3.97	(0.03)	0.05	(0.01)	0.64	(0.02)	0.07
	ILS	259	3.88	(0.03)	0.05	(0.01)	0.74	(0.02)	0.07
	LS	259	3.83	(0.03)	0.06	(0.01)	0.59	(0.02)	0.09
	ACT	259	3.84	(0.03)	0.02	(0.01)	0.60	(0.02)	0.04
	QU	259	3.57	(0.03)	0.05	(0.02)	0.51	(0.02)	0.10
	FEED	259	3.69	(0.03)	0.04	(0.01)	0.59	(0.02)	0.06
	TH	259	3.43	(0.03)	0.04	(0.01)	0.53	(0.02)	0.07
	MON	259	3.69	(0.03)	0.06	(0.02)	0.60	(0.02)	0.09
	CK	259	4.36	(0.03)	0.06	(0.01)	0.50	(0.02)	0.10
	KS	259	3.67	(0.03)	0.06	(0.02)	0.63	(0.02)	0.08
School Services Personnel **(ICC: 0.13)**	SO	1,562	4.07	(0.03)	0.06	(0.01)	0.57	(0.01)	0.09
	MS	1,562	4.25	(0.02)	0.05	(0.01)	0.51	(0.01)	0.08
	ACT	1,562	4.06	(0.02)	0.04	(0.01)	0.57	(0.01)	0.06
	PS	1,562	3.98	(0.03)	0.07	(0.01)	0.59	(0.01)	0.10
	KS	1,562	4.17	(0.02)	0.04	(0.01)	0.53	(0.01)	0.07
	DPS	1,562	4.19	(0.02)	0.05	(0.01)	0.55	(0.01)	0.08
	SS	1,562	4.02	(0.02)	0.05	(0.01)	0.54	(0.01)	0.08
	COM	1,562	4.09	(0.02)	0.05	(0.01)	0.56	(0.01)	0.08
	CON	1,562	4.12	(0.03)	0.06	(0.01)	0.55	(0.01)	0.10
	DEV	1,562	4.18	(0.03)	0.06	(0.01)	0.59	(0.01)	0.09
	OS	1,562	3.98	(0.02)	0.05	(0.01)	0.59	(0.01)	0.07

Note: Mixed-effects regression (intercept only) with district specified as the random effect. Values in parenthesis represent standard errors. Highlighted values represent domains/observation type with high intraclass correlation (≥0.10), indicating differences by districts.

Table A3. District-level Variability for Planning: Mixed-effects Regression

Observation	Indicator	n	Fixed Effect		Random Effect				Intraclass Correlation
			District (Intercept)		Var(District)		Var(Residual)		
General Educator (ICC: 0.13)	IP	6,057	3.89	(0.03)	0.08	(0.01)	0.62	(0.00)	0.11
	SW	6,057	3.63	(0.02)	0.07	(0.01)	0.62	(0.00)	0.10
	AS	6,057	3.46	(0.02)	0.07	(0.01)	0.63	(0.00)	0.10
Library Media Specialist (ICC: 0.06)	MCM	136	4.26	(0.03)	0.04	(0.01)	0.57	(0.02)	0.06
	MCR	136	4.17	(0.03)	0.03	(0.01)	0.56	(0.02)	0.05
	MCC	136	4.04	(0.03)	0.03	(0.01)	0.70	(0.03)	0.04
School Services Personnel (ICC: 0.13)	SOW	882	4.22	(0.03)	0.06	(0.01)	0.52	(0.01)	0.11
	AWP	882	4.13	(0.03)	0.05	(0.01)	0.61	(0.01)	0.08
	EVAL	882	3.90	(0.03)	0.08	(0.02)	0.61	(0.01)	0.11

Note: Mixed-effects regression (intercept only) with district specified as the random effect. Values in parenthesis represent standard errors. Highlighted values represent domains/observation type with high intraclass correlation (≥0.10), indicating differences by districts.

Table A4. District-level Variability for Environment: Mixed-effects Regression

Observation	Indicator	*n*	Fixed Effect		Random Effect				Intraclass Correlation
			District (Intercept)		Var(District)		Var(Residual)		
General Educator (ICC: 0.10)	EX	6,111	3.90	(0.02)	0.05	(0.01)	0.66	(0.00)	0.07
	MSB	6,114	4.02	(0.02)	0.05	(0.01)	0.74	(0.00)	0.06
	ENV	6,111	4.08	(0.02)	0.06	(0.01)	0.67	(0.00)	0.08
	RC	6,114	4.19	(0.02)	0.06	(0.01)	0.60	(0.00)	0.09
Library Media Specialist (ICC: 0.07)	EX	141	4.06	(0.03)	0.05	(0.02)	0.55	(0.02)	0.08
	MSB	141	4.12	(0.03)	0.02	(0.01)	0.60	(0.03)	0.04
	ENV	141	4.29	(0.03)	0.02	(0.01)	0.56	(0.02)	0.03
	RC	141	4.34	(0.03)	0.04	(0.01)	0.48	(0.02)	0.08
School Services Personnel (ICC: 0.14)	EX	869	4.22	(0.03)	0.05	(0.01)	0.50	(0.01)	0.09
	MSB	869	4.18	(0.03)	0.06	(0.01)	0.56	(0.01)	0.10
	ENV	869	4.36	(0.03)	0.06	(0.01)	0.50	(0.01)	0.10
	RC	869	4.47	(0.03)	0.05	(0.01)	0.41	(0.01)	0.11

Note: Mixed-effects regression (intercept only) with district specified as the random effect. Values in parenthesis represent standard errors. Highlighted values represent domains/observation type with high intraclass correlation (≥0.10), indicating differences by districts.

Table A5. District-level Variability for Professionalism: Mixed-effects regression

Observation	Indicator	*n*	Fixed Effect		Random Effect				Intraclass Correlation
			District (Intercept)		Var(District)		Var(Residual)		
Professionalism (ICC: 0.18)	PROF01	5,679	4.13	(0.02)	0.07	(0.01)	0.62	(0.00)	0.11
	PROF02	5,679	4.05	(0.03)	0.08	(0.01)	0.60	(0.00)	0.12
	PROF03	5,679	3.89	(0.03)	0.09	(0.01)	0.62	(0.00)	0.12
	PROF04	5,679	3.98	(0.03)	0.10	(0.01)	0.62	(0.00)	0.13
	PROF05	5,679	4.04	(0.03)	0.09	(0.01)	0.62	(0.00)	0.13
	PROF06	5,679	3.99	(0.03)	0.07	(0.01)	0.68	(0.00)	0.10
	PROF07	5,679	4.20	(0.02)	0.07	(0.01)	0.68	(0.00)	0.09
	PROF08	5,679	4.27	(0.03)	0.11	(0.01)	0.67	(0.00)	0.14
	PROF09	5,679	4.22	(0.03)	0.11	(0.02)	0.60	(0.00)	0.16
	PROF10	5,679	3.96	(0.03)	0.08	(0.01)	0.72	(0.00)	0.10

Note: Mixed-effects regression (intercept only) with district specified as the random effect. Values in parenthesis represent standard errors. Highlighted values represent domains/observation type with high intraclass correlation (≥0.10), indicating differences by districts.

Table A6. Detailed Revisions to Instruction Rubric

		Significantly Above Expectations (5)	At Expectations (3)	Significantly Below Expectations (1)
Standards and Objectives	1	Learning objectives and their purpose are clearly stated and consistently referenced throughout the lesson.	Learning objectives and their purpose are stated and sometimes referenced in the lesson.	Learning objectives and their purpose are seldom stated and/or referenced during the lesson.
	2	Expectations for student performance are clearly and consistently stated and/or modeled.	Expectations for student performance are stated and/or modeled.	Expectations for student performance are vague.

		Significantly Above Expectations (5)	At Expectations (3)	Significantly Below Expectations (1)
Presenting Instructional Content	3	Presentation of content always previews the organization of the lesson.	Presentation of content most of the time previews preview the organization of the lesson.	Presentation of content rarely previews the organization of the lesson.
	4	Presentation of content always includes examples, illustrations, analogies, and labels for new concepts and ideas.	Presentation of content most of the time includes examples, illustrations, analogies, and labels for new concepts and ideas.	Presentation of content rarely includes examples, illustrations, analogies, and/or labels for new concepts and ideas.
	5	Presentation of content always includes content-focused communication.	Presentation of content most of the time includes content-focused communication.	Presentation of content rarely includes content-focused communication.
	6	Presentation of content always includes logical sequencing and segmenting.	Presentation of content most of the time includes logical sequencing and segmenting.	Presentation of content rarely includes logical sequencing and segmenting.

		Significantly Above Expectations (5)	At Expectations (3)	Significantly Below Expectations (1)
Lesson Structure and Pacing	7	Lesson plan is strongly reflected in the observed lesson.	Lesson plan is somewhat reflected in the observed lesson.	Lesson plan is minimally reflected in the observed lesson.
	8	Pacing provides many opportunities for individual students who progress at different learning rates.	Pacing sometimes provides opportunities for students who progress at different learning rates.	Pacing rarely provides opportunities for students who progress at different learning rates.
	9	Routines for distributing materials are highly efficient.	Routines for distributing materials are efficient.	Routines for distributing materials are inefficient.
	10	No instructional time is lost during transitions.	Little instructional time is lost during transitions.	Considerable time is lost during transitions.

Table A6. Detailed Revisions to Instruction Rubric (cont.)

Activities and Materials	11	Activities and materials consistently align to the lesson objectives.	Activities and materials sometimes support the lesson objectives.	Activities and materials rarely support the lesson objectives.
	12	Activities and materials are consistently cognitively challenging, as demonstrated by a combination of: • providing learning activities that build on prior knowledge, • learning is done by the learner, • offering different ways of demonstrating knowledge, • providing opportunities for student-to student interaction, • differentiation with respect to ability, interest, background, and learning style.	Activities and materials are sometimes cognitively challenging, as demonstrated by a combination of: • providing learning activities that build on prior knowledge, • learning is done by the learner, • offering different ways of demonstrating knowledge, • providing opportunities for student-to student interaction, • differentiation with respect to ability, interest, background, and learning style.	Activities and materials are rarely cognitively challenging, as demonstrated by a combination of: • providing learning activities that build on prior knowledge, • learning is done by the learner, • offering different ways of demonstrating knowledge, • providing opportunities for student-to student interaction, • differentiation with respect to ability, interest, background, and learning style.
	13	Activities and materials consistently incorporate multimedia, technology, and resources beyond the classroom in ways that support lesson objectives.	Activities and materials sometimes incorporate multimedia, technology, and resources beyond the classroom in ways that support lesson objectives.	Activities and materials rarely incorporate multimedia, technology, and resources beyond the classroom in ways that support lesson objectives.

		Significantly Above Expectations (5)	At Expectations (3)	Significantly Below Expectations (1)
Questioning	14	Teacher questions are varied and of high quality, providing a balanced mix of question types: • knowledge and comprehension; • application and analysis; and • creation and evaluation.	Teacher questions are varied and of high quality providing for some, but not all, question types: • knowledge and comprehension; • application and analysis; and • creation and evaluation.	Teacher questions are inconsistent in quality and include few question types: • knowledge and comprehension; • application and analysis; and • creation and evaluation.
	15	Teacher questions are consistently aligned with lesson objectives.	Teacher questions are sometimes aligned with lesson objectives.	Teacher questions are rarely aligned with lesson objectives.

Table A6. Detailed Revisions to Instruction Rubric (cont.)

	16	Questioning consistently engages all students, through: • consistently providing 3-5 seconds wait time for students to prepare to respond • varying response strategies (e.g. random, patterned, choral responses, group or individual answers from a balance of students based on ability, gender, and background).	Questioning sometimes engages all students, through: • consistently providing 3-5 seconds wait time for students to prepare to respond • varying response strategies (e.g. random, patterned, choral responses, group or individual answers from a balance of students based on ability, gender, and background).	Questioning rarely engages all students, through: • consistently providing 3-5 seconds wait time for students to prepare to respond • varying response strategies (e.g. random, patterned, choral responses, group or individual answers from a balance of students based on ability, gender, and background).
	17	Students sometimes generate questions that lead to further inquiry and self-directed learning.	Students rarely generate questions that lead to further inquiry and self-directed learning.	Students never questions that lead to further inquiry and self-directed learning.

		Significantly Above Expectations (5)	At Expectations (3)	Significantly Below Expectations (1)
Academic Feedback	18	Feedback is consistently specific and promotes progress towards lesson objectives.	Feedback is sometimes specific and promotes progress towards lesson objectives.	Feedback is rarely specific or promotes progress towards lesson objectives.
	19	Teacher consistently prompts student thinking, assesses each student's progress, and provides individual feedback.	Teacher sometimes prompts student thinking, assesses each student's progress, and provides individual feedback.	Teacher rarely prompts student thinking, assesses each student's progress, or provides individual feedback.
	20	Feedback from students is consistently used to monitor and adjust instruction; or instruction does not need to be adjusted.	Feedback from students is sometimes used to monitor and adjust instruction.	Feedback from students is rarely used to monitor and adjust instruction.
	21	Teacher sometimes engages students in giving specific and high-quality feedback to one another.	Teacher rarely engages students in giving specific and high-quality feedback to one another.	Teacher never engages students in giving specific and high-quality feedback to one another.

Table A7. Detailed Revisions to Planning Rubric

Instructional Plans		Significantly Above Expectations (5)	Above Expectations (3)	Significantly Below Expectations (1)
Instructional Plans	1	Instructional plans include explicit and clear goals aligned to state content standards.	Instructional plans include goals aligned to state content standards.	Instructional plans include no goals or goals not aligned to state content standards.
	2	Instructional plans include activities, materials, and assessments that are aligned to instructional goals.	Instructional plans include activities, materials, and assessments that are aligned to instructional goals.	Instructional plans include activities, materials, and assessments that are rarely aligned to instructional goals.
	3	Instructional plans consistently include activities, materials, and assessments that are sequenced from basic to complex.	Instructional plans sometimes include activities, materials, and assessments that are sequenced from basic to complex.	Instructional plans rarely include activities, materials, and assessments that are logically sequenced.
	4	Instructional plans consistently include activities, materials, and assessments that build on prior student knowledge, student interest, and integrate other disciplines.	Instructional plans sometimes include activities, materials, and assessments that build on prior student knowledge, student interest, and integrate other disciplines.	Instructional plans rarely include activities, materials, and assessments that build on prior student knowledge, student interest, or integrate other disciplines.
	5	Instructional plans are consistently based on formative and summative school and individual student achievement data.	Instructional plans are sometimes based on formative and summative school and individual student achievement data.	Instructional plans are rarely based on formative and summative school and individual student achievement data.
	6	The instructional group composition is consistently varied (e.g., race, gender, ability, age) to best accomplish the goals of the lesson.	The instructional group composition is sometimes varied (e.g., race, gender, ability, age) to best accomplish the goals of the lesson.	The instructional group composition is rarely varied (e.g., race, gender, ability, age) to best accomplish the goals of the lesson.
	7	Instructional plans consistently provide opportunities to accommodate individual student needs.	Instructional plans provide some opportunities to accommodate individual student needs.	Instructional plans rarely provide opportunities to accommodate individual student needs.

Table A7. Detailed Revisions to Planning Rubric (cont.)

Teacher Pedagogical Content Knowledge[8]	8	Planning consistently reflects knowledge of how to structure and represent academic content for direct teaching to students.	Planning sometimes reflects knowledge of how to structure and represent academic content for direct teaching to students.	Planning rarely reflects knowledge of how to structure and represent academic content for direct teaching to students.
	9	Planning consistently reflects knowledge of the common conceptions, misconceptions, and difficulties that students encounter when learning particular content.	Planning sometimes reflects knowledge of the common conceptions, misconceptions, and difficulties that students encounter when learning particular content.	Planning rarely reflects knowledge of the common conceptions, misconceptions, or difficulties that students encounter when learning particular content.
	10	Planning consistently reflects knowledge of the specific teaching strategies that can be used to address students' learning needs in the particular classroom circumstances.	Planning sometimes reflects knowledge of the specific teaching strategies that can be used to address students' learning needs in the particular classroom circumstances.	Planning rarely reflects knowledge of the specific teaching strategies that can be used to address students' learning needs in the particular classroom circumstances.

Student Work	11	Assignments consistently require students to create, evaluate, analyze, apply, and understand information rather than only remember it.	Assignments sometimes require students to create, evaluate, analyze, apply, and understand information rather than only remember it.	Assignments rarely require students to create, evaluate, analyze, apply, and understand information rather than only remember it.
	12	Assignments consistently require students to draw conclusions, make generalizations, and produce arguments that are supported through extended writing.	Assignments consistently require students to draw conclusions, make generalizations, and produce arguments that are supported through extended writing.	Assignments consistently require students to draw conclusions, make generalizations, and produce arguments that are supported through extended writing.
	13	Assignments consistently require students to connect what they are learning to experiences, observations, feelings, or situations significant in their daily lives both inside and outside of school.	Assignments consistently require students to connect what they are learning to experiences, observations, feelings, or situations significant in their daily lives both inside and outside of school.	Assignments consistently require students to connect what they are learning to experiences, observations, feelings, or situations significant in their daily lives both inside and outside of school.

---

[8] This breakdown of Lee Shulman's concept of Pedagogical Content Knowledge is taken from Rowan et al. (2001). Measuring teachers' pedagogical content knowledge in surveys: An exploratory study. *Consortium for Policy Research in Education.*

Table A7. Detailed Revisions to Planning Rubric (cont.)

		Significantly Above Expectations (5)	At Expectations (3)	Significantly Below Expectations (1)
Assessment	14	Assessment plans are aligned with state content standards.	Assessment plans are somewhat aligned with state content standards.	Assessment plans are rarely aligned with state content standards.
	15	Assessment plans have clear, measureable criteria for progressing towards the standards.	Assessment plans have measureable criteria for progressing towards the standards.	Assessment plans do not have measureable criteria for progressing towards the standards.
	16	Assessment plans provide multiple ways to measure student performance (e.g., in the form of a project, experiment, presentation, essay, short answer, or multiple choice test).	Assessment plans provide a few ways to measure student performance (e.g., in the form of a project, experiment, presentation, essay, short answer, or multiple choice test).	Assessment plans does not provide more than one way to measure student performance (e.g., in the form of a project, experiment, presentation, essay, short answer, or multiple choice test).
	17	Assessment plans are portfolio-based with clear illustrations of student progress toward state content standards.	Assessment plans are portfolio based with some illustration of student progress toward state content standards.	Assessment plans are not portfolio based.
	18	Assessment plans include descriptions of how assessment results will be used to inform future instruction.	Assessment plans include some description of how the assessment results will be used to inform future instruction.	Assessment plans do not provide any description of how the assessment results will be used to inform future instruction.

Table A8. Detailed Revisions to Environment Rubric

		Significantly Above Expectations (5)	At Expectations (3)	Significantly Below Expectations (1)
Expectations	1	Teacher clearly sets high academic expectations for every student.	Teacher sets high academic expectations for every student.	Teacher does not clearly set high expectations for every student.
	2	Teacher consistently encourages students to make multiple attempts, possibly make mistakes, and to learn from them.	Teacher sometimes encourages students to make multiple attempts, possibly make mistakes, and to learn from them.	Teacher rarely encourages students to make multiple attempts, possibly make mistakes, and to learn from them.
Managing Student Behavior	3	Students are consistently well-behaved and on task in individual, group, and whole class settings.	Students are mostly well-behaved and on task, some minor learning disruptions may occur.	Students are not well-behaved and are often off task.
	4	Teacher and students communicate clear rules for learning and behavior; or there is no misbehavior.	Teacher communicates clear rules for learning and behavior.	Teacher does not clearly communicate any rules for learning and behavior.
	5	The teacher uses several techniques, such as social approval, contingent activities, and consequences to maintain appropriate student behavior.	The teacher uses some techniques, such as social approval, contingent activities, and consequences to maintain appropriate student behavior.	The teacher uses few techniques to maintain appropriate student behavior.
	6	The teacher attends to disruptions in a respectful and timely manner and with complete success; or there are no disruptions.	Teacher somewhat attends to disruption in a timely and respectful manner and with some success.	Teacher does not attend to disruptions, or attention to disruption is not timely, respectful or successful.
The Classroom's Physical Environment	7	The classroom is well organized with supplies, equipment, and resources easily and readily accessible.	The classroom is organized with supplies, equipment, and resources accessible.	The classroom is not well organized with supplies, equipment, and resources difficult to access.
	8	The classroom displays recent student work, with a clear purpose.	The classroom displays recent student work.	The classroom does not display recent student work.
	9	The classroom is always arranged to promote individual and group learning.	The classroom is sometimes arranged to promote individual and group learning.	The classroom is never arranged to promote group learning.

		Significantly Above Expectations (5)	At Expectations (3)	Significantly Below Expectations (1)
Respectful Culture	10	Teacher-student interactions consistently demonstrate caring and respect for one another.	Teacher-student interactions demonstrate caring and respect for one another.	Teacher-student interactions rarely demonstrate caring and respect for one another.
	11	Students consistently exhibit caring and respect for one another.	Students generally exhibit caring and respect for one another.	Students rarely exhibit caring and respect for one another.
	12	Teacher elicits and is receptive to the interests and opinions of all students.	Teacher sometimes elicits and is receptive to the interests and opinions of students.	Teacher is not receptive to interests and opinions of students.

Table A9. Detailed Revision of Professional Rubric

		Significantly Above Expectations (5)	At Expectations (3)	Significantly Below Expectations (1)
Professional Growth and Learning	1	Demonstrates extensive engagement in the evaluation process by seeking feedback about their practice from multiple sources: observations, professional conversations with evaluators and peers, assessment and/or survey data, and school/district community members.	Demonstrates engagement in the evaluation process by seeking feedback about their practice from multiple sources: observations, professional conversations with evaluators and peers, assessment and/or survey data, school/district community members.	Demonstrates limited engagement in the evaluation process by seeking feedback about their practice from multiple sources: observations, professional conversations with evaluators and peers, assessment and/or survey data, school/district community members.
	2	Reflects on all evidence of practice and aligns that evidence accurately with the TEAM rubric to appropriately determine areas of strength and areas for growth.	Reflects on some evidence of practice and with some precision aligns that evidence accurately with the TEAM rubric to appropriately determine areas of strength and areas for growth.	Rarely reflects on evidence of practice and does not align that evidence accurately with the TEAM rubric to appropriately determine areas of strength and areas for growth.
	3	Consistently collaborates with evaluators and peers, reflecting on their feedback on practice.	Sometimes collaborates with evaluators and peers, reflecting on their feedback on practice.	Rarely collaborates with evaluators and peers, reflecting on their feedback on practice.

		Significantly Above Expectations (5)	At Expectations (3)	Significantly Below Expectations (1)
School and Community Involvement	4	Consistently organizes and leads school activities and events that positively impact school results and culture.	Consistently supports and contributes to school activities and events.	Rarely supports school activities and events.
	5	Always adheres to school and district personnel policies.	Consistently adheres to school and district personnel policies.	Sometimes adheres to school and district personnel policies.
	6	Consistently collaborates with peers to contribute to a safe and orderly learning environment	Sometimes collaborates with peers to contribute to a safe and orderly learning environment.	Rarely collaborates with peers to contribute to a safe and orderly learning environment.

Leadership	7	Actively and consistently contributes to the school community by assisting and/or mentoring others, as characterized by combinations of the following: • Collaborative planning with subject and/or grade level teams • Actively leading in a Professional Learning Community • Coaching/mentoring • Supervising clinical experiences • Leading data driven professional learning opportunities.	Contributes to the school community by assisting and/or mentoring others, including but not limited to the following: • Collaborative planning with subject and/or grade level teams • Actively participating in a Professional Learning Community • Coaching/mentoring • Supervising clinical experiences • Leading data driven professional learning opportunities.	Inconsistently contributes to the school community by assisting and/or mentoring others, including but not limited to the following: • Collaborative planning with subject and/or grade level teams • Actively participating in a Professional Learning Community • Coaching/mentoring • Supervising clinical experiences • Leading data driven professional learning opportunities.